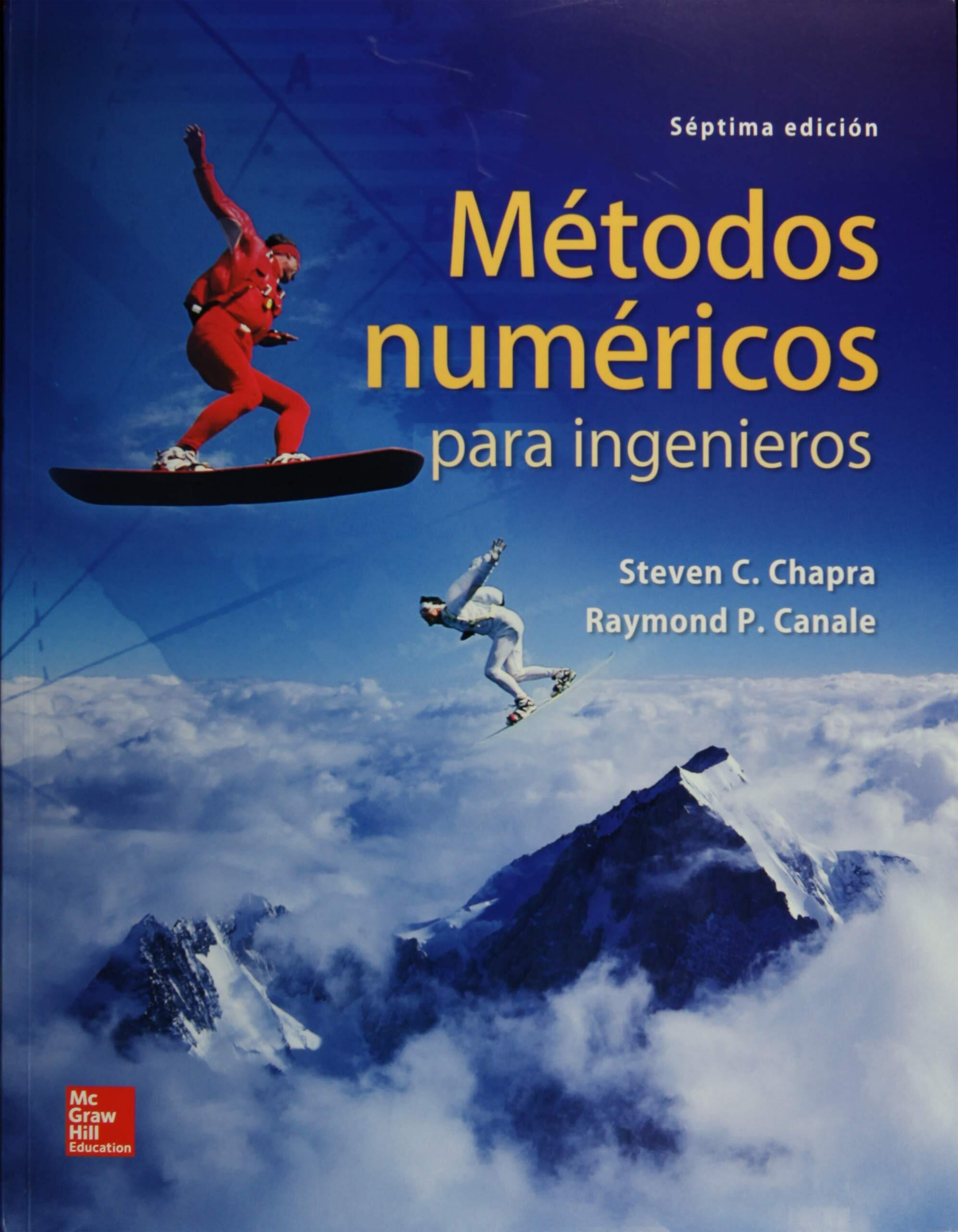

Séptima edición
Métodos numéricos
para ingenieros
Steven C. Chapra
Raymond P. Canale
Mc Graw Hill Education

Métodos numéricos

para ingenieros

Métodos numéricos
para ingenieros

Séptima edición

Steven C. Chapra
Tufts University

Raymond P. Canale
University of Michigan

Revisión técnica

Reyna Susana García Ruiz
Universidad Iberoamericana, Ciudad de México

Víctor Hugo Ibarra Mercado
Universidad Anáhuac, México Norte

Enrique Muñoz Díaz
Instituto Tecnológico y de Estudios Superiores de Monterrey,
Campus Ciudad de México

Guillermo Evangelista Benites
Universidad Nacional de Trujillo, Perú



MÉXICO • AUCKLAND • BOGOTÁ • BUENOS AIRES • GUATEMALA • LONDRES
MADRID • MILÁN • MONTREAL • NUEVA DELHI • NUEVA YORK • SAN FRANCISCO
SAN JUAN • SANTIAGO • SAO PAULO • SIDNEY • SINGAPUR • ST. LOUIS • TORONTO

Directora de desarrollo de contenido editorial y digital: Patricia Ledezma Llaca
Coordinador sponsor: Jesús Mares Chacón
Coordinadora editorial: Marcela I. Rocha Martínez
Editora de desarrollo: Ana L. Delgado Rodríguez
Supervisor de producción: Zeferino García García
Traducción: Sergio M. Sarmiento Ortega

Métodos numéricos
para ingenieros

Séptima edición



Edificio Punta Santa Fe
Prolongación Paseo de la Reforma 1015, Torre A,
Piso 16, Colonia Desarrollo Santa Fe,
Delegación Álvaro Obregón,
C.P. 01376, México, D. F.
Miembro de la Cámara Nacional de la Industria Editorial Mexicana, Reg. Núm. 736

ISBN: 978-607-15-1294-9
ISBN (de la sexta edición): 978-607-15-0499-9

Traducido de la séptima edición de *Numerical Methods for Engineers* by Steven C. Chapra and Raymond P. Canale.

JEC 05/15

1234567890 2346789015

Impreso en México *Printed in Mexico*

 *A*

Margaret y Gabriel Chapra

Helen y Chester Canale

CONTENIDO

PARTE TRES ECUACIONES ALGEBRAICAS LINEALES 181

PARTE CUATRO OPTIMIZACIÓN 273

PREFACIO

Han pasado veinte años desde que se publicó por primera vez este libro. Durante ese tiempo, nuestro escepticismo acerca de que los métodos numéricos y las computadoras tendrían un papel prominente en el currículo de la ingeniería —particularmente en sus etapas tempranas— ha sido rebasado por mucho. Hoy día, muchas universidades ofrecen cursos para estudiantes de nuevo ingreso, de segundo año e intermedios, tanto de introducción a la computación como de métodos numéricos. Además, muchos de nuestros colegas integran problemas orientados a la computación con otros cursos, en todos los niveles del currículo. Así, esta nueva edición aún se basa en la premisa fundamental de que debe darse a los estudiantes de ingeniería una introducción profunda y temprana a los métodos numéricos. En consecuencia, aunque la nueva edición expande sus alcances, tratamos de mantener muchas de las características que hicieron accesible la primera edición, para estudiantes principiantes como avanzados:

- **Orientado a problemas.** Los estudiantes de ingeniería aprenden mejor cuando están motivados por la solución de problemas, lo cual es especialmente cierto en el caso de las matemáticas y de la computación. Por tal razón, presentamos los métodos numéricos desde la perspectiva de la solución de problemas.
- **Pedagogía enfocada en el estudiante.** Hemos presentado varios detalles para lograr que este texto sea tan accesible para el estudiante como sea posible. Estos detalles comprenden la organización general, el uso de introducciones y epílogos para consolidar los temas principales, así como un amplio uso de ejemplos desarrollados y estudios de casos de las áreas principales de la ingeniería. Hemos puesto especial cuidado en que nuestras explicaciones sean claras y en que tengan una orientación práctica.
- **Herramientas de computación.** Capacitamos a nuestros estudiantes ayudándoles a utilizar las herramientas numéricas tipo "apunte y dispare" para resolución de problemas que están incluidas en programas como Excel, MATLAB y Mathcad. Sin embargo, también se muestra a los estudiantes cómo desarrollar programas sencillos y bien estructurados para extender las capacidades básicas de dichos entornos. Este conocimiento incluye lenguajes de programación estándar tales como Visual Basic, Fortran 90 y C/C++. Creemos que el abandono actual de la programación de computadora representa algo así como una "salida fácil" en los planes de estudios de ingeniería. A fin de cuentas, en la medida que los ingenieros no se conformen con herramientas limitadas, tendrán que escribir sus códigos. Sólo entonces se podrán llamar "macros" o "archivos M". Este libro está diseñado para darles el poder de hacer eso.

Más allá de estos cinco principios originales, la séptima edición tiene conjuntos de problemas nuevos y expandidos. La mayoría de los problemas se han modificado de manera que tengan soluciones numéricas diferentes de las ediciones anteriores. Además, se han incluido diversos problemas nuevos.

La séptima edición también incluye **Connect® de McGraw-Hill**. Esta herramienta de administración de tareas en línea permite la asignación de problemas algorítmicos para tareas en casa, exámenes y pruebas. Conecta a los estudiantes con las herramientas y los recursos que necesitan para lograr el éxito.

Como siempre, nuestra intención principal al escribir este libro es dar a los estudiantes una introducción a los métodos numéricos. Creemos que los estudiantes motivados que disfrutan de los métodos numéricos, las computadoras y las matemáticas serán, a final de cuentas, mejores

ingenieros. Si nuestro texto fomenta el entusiasmo por estos temas, consideraremos que nuestros esfuerzos han tenido buen éxito,

Agradecimientos. Nos gustaría agradecer a nuestros amigos de McGraw-Hill. Especialmente, Lorraine Buczek y Bill Stenquist, quienes brindaron una atmósfera positiva de apoyo para crear esta edición. Como de costumbre, Beatrice Sussman hizo un trabajo magistral en el cotejo del manuscrito, y Arpana Kumari de Aptara también realizó un trabajo excelente en la fase final de producción del libro. Como en ediciones pasadas, David Clough (Universidad de Colorado), Mike Gustafson (Universidad de Duke) y Jerry Stedinger (Universidad Cornell) compartieron generosamente sus observaciones penetrantes y sus sugerencias. También hicieron sugerencias útiles Bill Philpot (Universidad Cornell), Jim Guilkey (Universidad de Utah), Dong-Il Seo (Universidad Nacional Chungnam de Corea), Niall Broekhuizen (NIWA, Nueva Zelanda) y Raymundo Cordero y Karim Muci (ITESM, México). La edición actual también ha aprovechado las revisiones y sugerencias de los siguientes colegas:

Betty Barr, University of Houston
Jalal Behzadi, Shahid Chamran University
Jordan Berg, Texas Tech University
Jacob Bishop, Utah State University
Estelle M. Eke, California State University, Sacramento
Yazan A. Hussain, Jordan University of Science & Technology
Yogesh Jaluria, Rutgers University
S. Graham Kelly, The University of Akron
Subha Kumpaty, Milwaukee School of Engineering
Eckart Meiburg, University of California-Santa Barbara
Prashant Mhaskar, McMaster University
Luke Olson, University of Illinois at Urbana-Champaign
Richard Pates Jr., Old Dominion University
Joseph H. Pierluissi, University of Texas at El Paso
Juan Perán, Universidad Nacional de Educación a Distancia (UNED)
Scott A. Socolofsky, Texas A&M University

Se debe subrayar que, aunque recibimos consejos útiles de las personas arriba mencionadas, somos responsables de cualquier inexactitud o error que usted pudiera detectar en esta edición, Por favor póngase en contacto con Steve Chapra por correo electrónico si detectara cualquier error en esta edición.

Finalmente, nos gustaría dar las gracias a nuestras familias, a nuestros amigos y a nuestros estudiantes por su constante paciencia y apoyo. En especial a Cynthia Chapra, Danielle Husley y Claire Canale, quienes están siempre presentes para brindar entendimiento, perspectiva y amor.

Steve C. Chapra
Medford, Massachusetts

Raymond P. Canale
Lake Leelanau, Michigan

ACERCA DE LOS AUTORES

Steven Chapra es profesor en el Departamento de Ingeniería Civil y Ambiental de la Universidad de Tufts. Entre sus obras publicadas se encuentran *Surface Water-Quality Modeling* e *Introduction to Computing for Engineers*.

El doctor Chapra obtuvo el grado de ingeniero por las universidades de Manhattan y de Michigan. Antes de incorporarse a la Tufts, trabajó para la Agencia de Protección Ambiental y la Administración Nacional del Océano y la Atmósfera, fue profesor asociado en las universidades de Texas A&M y de Colorado. En general, sus investigaciones están relacionadas con la modelación de la calidad del agua superficial y la aplicación de computación avanzada en la ingeniería ambiental.

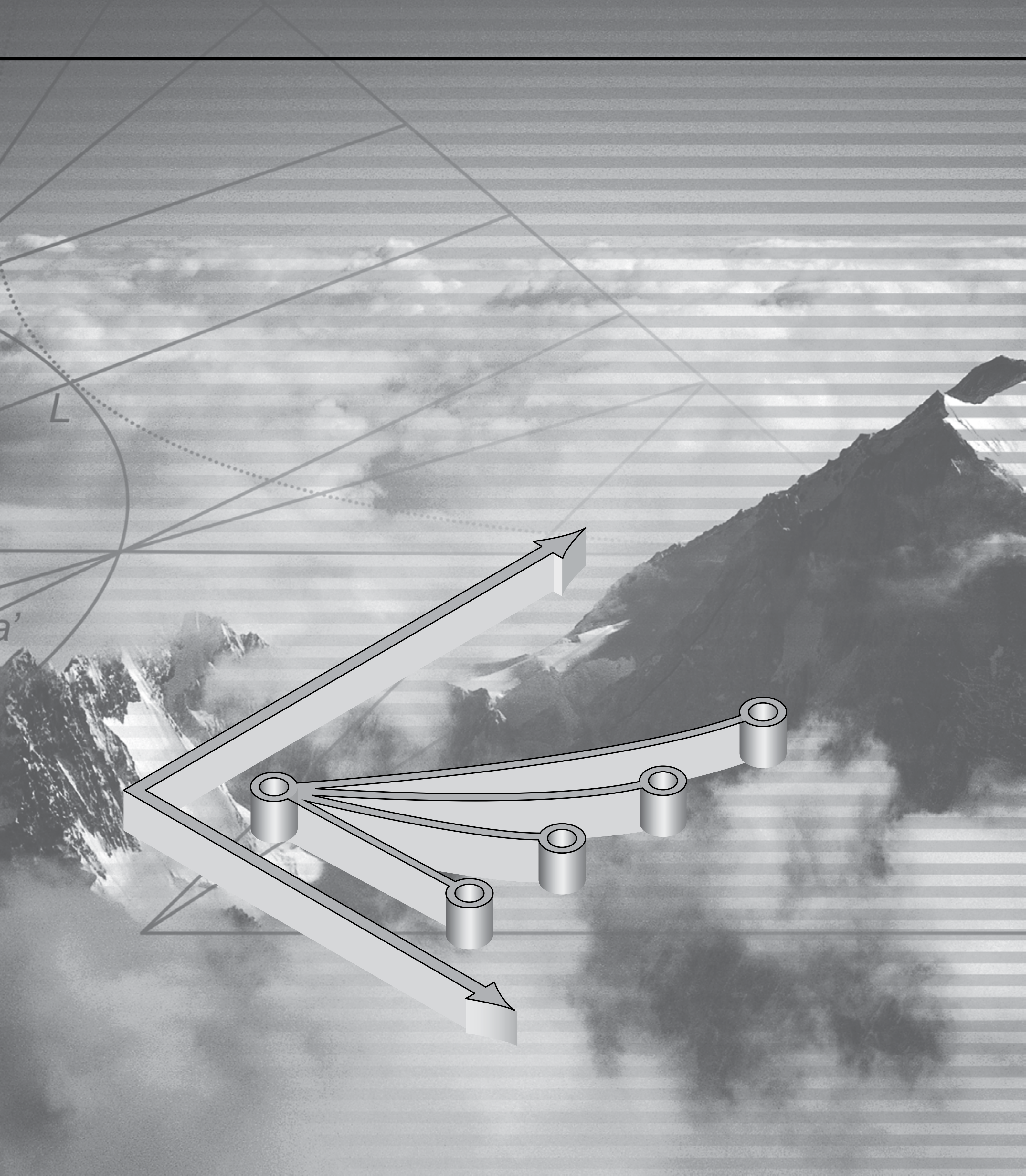
También ha recibido gran cantidad de reconocimientos por sus destacadas contribuciones académicas, incluyendo la medalla Rudolph Hering (ASCE en 1993) y el premio al autor distinguido Meriam-Wiley (por parte de la Sociedad Estadounidense para la Educación en Ingeniería). Se le ha reconocido como profesor emérito en las facultades de ingeniería de las universidades de Texas A&M, de Colorado y Tufts.

Raymond P. Canale es profesor emérito de la Universidad de Michigan. En sus más de 20 años como docente universitario ha impartido numerosos cursos en las áreas de computación, métodos numéricos e ingeniería ambiental. También ha dirigido extensos programas de investigación en el área de modelación matemática y por computadora de ecosistemas acuáticos. Es autor y coautor de varios libros, ha publicado más de 100 artículos e informes científicos. También ha diseñado y desarrollado software para computadoras personales, con la finalidad de facilitar la educación en ingeniería y la solución de problemas en ingeniería. Ha recibido el premio al autor distinguido Meriam-Wiley de la Sociedad Americana para la Educación en Ingeniería por sus libros y el software desarrollado, así como otros reconocimientos por sus publicaciones técnicas.

Actualmente, el profesor Canale se dedica a resolver problemas de aplicación, trabajando como consultor y perito en empresas de ingeniería, en la industria e instituciones gubernamentales.

Métodos numéricos
para ingenieros

PARTE UNO



MODELOS, COMPUTADORAS Y ANÁLISIS DEL ERROR

PT1.1 MOTIVACIÓN

Los métodos numéricos constituyen técnicas mediante las cuales es posible formular problemas matemáticos, de tal forma que puedan resolverse utilizando operaciones aritméticas. Aunque existen muchos tipos de métodos numéricos, éstos comparten una característica común: invariablemente requieren de un buen número de tediosos cálculos aritméticos. No es raro que con el desarrollo de computadoras digitales eficientes y rápidas, el papel de los métodos numéricos en la solución de problemas en ingeniería haya aumentado de forma considerable en los últimos años.

PT1.1.1 Métodos no computacionales

Además de proporcionar un aumento en la potencia de cálculo, la disponibilidad creciente de las computadoras (en especial de las personales) y su asociación con los métodos numéricos han influido de manera muy significativa en el proceso de la solución actual de los problemas en ingeniería. Antes de la era de la computadora los ingenieros sólo contaban con tres métodos para la solución de problemas:

1. Las soluciones de algunos problemas fueron obtenidas usando métodos exactos o analíticos. Dichas soluciones resultaban útiles y proporcionaban una comprensión excelente del comportamiento de algunos sistemas. No obstante, las soluciones analíticas sólo pueden encontrarse para una clase limitada de problemas. Éstos incluyen aquellos que pueden aproximarse mediante modelos lineales y también aquellos que tienen una geometría simple y de baja dimensión. En consecuencia, las soluciones analíticas tienen un valor práctico limitado porque la mayoría de los problemas reales son no lineales, e implican formas y procesos complejos.
2. Las soluciones gráficas fueron usadas para caracterizar el comportamiento de los sistemas, usualmente gráficas o monogramas, las cuales tomaban la forma de gráficas o monogramas; aunque las técnicas gráficas se utilizan a menudo para resolver problemas complejos, los resultados no son muy precisos. Además, las soluciones gráficas (sin la ayuda de una computadora) son en extremo tediosas y difíciles de implementar. Finalmente, las técnicas gráficas están limitadas a los problemas que puedan describirse usando tres dimensiones o menos.
3. Para implementar los métodos numéricos manualmente se utilizaban calculadoras y reglas de cálculo. Aunque en teoría dichas aproximaciones deberían ser perfectamente adecuadas para resolver problemas complicados, en la práctica se presentan varias dificultades debido a que los cálculos manuales son lentos y tediosos. Además, los resultados no son consistentes, ya que surgen equivocaciones cuando se efectúan los numerosos cálculos de esta manera.

Antes del uso de la computadora se gastaba bastante energía en la técnica misma de solución, en lugar de usarla en la definición del problema y su interpretación (figura PT1.1*a*). Esta situación desafortunada se debía al tiempo y trabajo monótono que se requería para obtener resultados numéricos con técnicas que no utilizaban la computadora.

En la actualidad, las computadoras y los métodos numéricos ofrecen una alternativa para los cálculos complicados. Al usar el poder del cómputo se obtienen soluciones directamente, de esta manera se pueden aproximar los cálculos sin tener que recurrir a consideraciones de simplificación o a técnicas muy lentas. Aunque las soluciones analíticas aún son muy valiosas, tanto para resolver problemas como para brindar una mayor comprensión, los métodos numéricos representan opciones que aumentan, en forma considerable, la capacidad para enfrentar y resolver los problemas; como resultado, se dispone de más tiempo para aprovechar las habilidades creativas personales. En

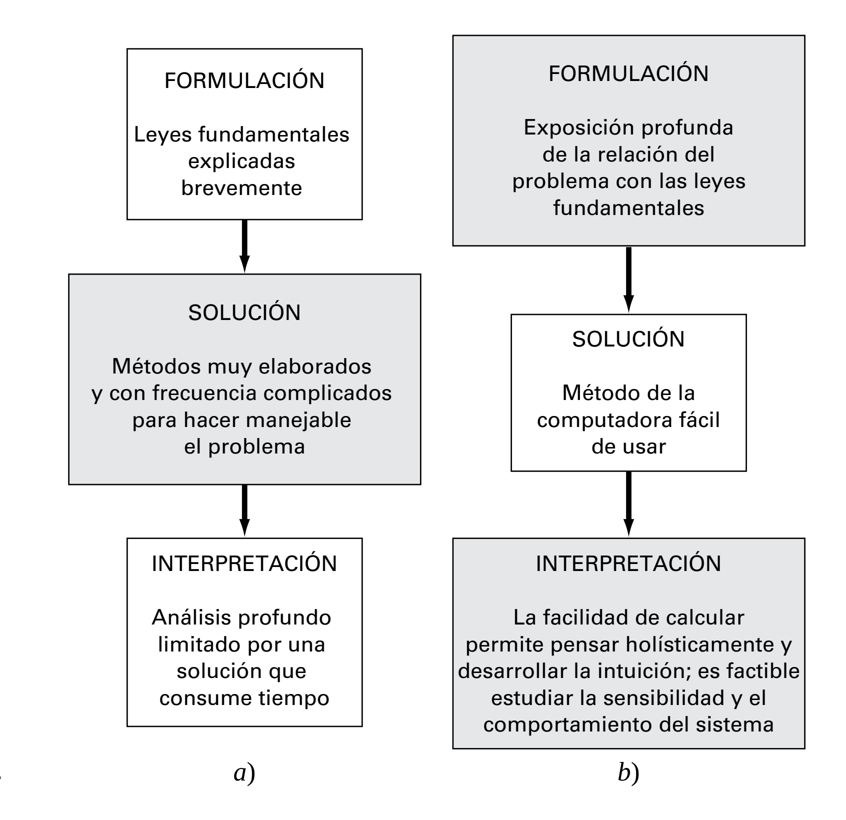


Figura PT1.1
Las tres fases en la solución de problemas en ingeniería en *a*) la era anterior a las computadoras y *b*) la era de las computadoras. Los tamaños de los recuadros indican el nivel de importancia que se presenta en cada fase. Las computadoras facilitan la implementación de técnicas de solución y, así, permiten un mayor interés sobre los aspectos creativos en la formulación de problemas y la interpretación de los resultados.

consecuencia, es posible dar más importancia a la formulación de un problema y a la interpretación de la solución, así como a su incorporación al sistema total, o conciencia "holística" (figura PT1.1*b*).

PT1.1.2 Los métodos numéricos y la práctica en ingeniería

Desde finales de la década de los cuarenta, la amplia disponibilidad de las computadoras digitales han llevado a una verdadera explosión en el uso y desarrollo de los métodos numéricos. Al principio, este crecimiento estaba limitado por el costo de acceso a las *grandes computadoras* (*mainframes*), por lo que muchos ingenieros seguían usando simples procedimientos analíticos en una buena parte de su trabajo. Vale la pena mencionar que la reciente evolución de computadoras personales de bajo costo ha permitido el acceso, de mucha gente, a las poderosas capacidades de cómputo. Además, existen diversas razones por las cuales se deben estudiar los métodos numéricos:

1. Los métodos numéricos son herramientas muy poderosas para la solución de problemas. Son capaces de manipular grandes sistemas de ecuaciones, manejar no linealidades y resolver geometrías complicadas, comunes en la práctica de la ingeniería y, a menudo, imposibles de resolver en forma analítica. Por lo tanto, aumentan la habilidad de quien los estudia para resolver problemas.
2. En el transcurso de su carrera, es posible que usted tenga la oportunidad de utilizar paquetes disponibles comercialmente, o programas "en caja negra" que contengan métodos numéricos. El uso eficiente de estos programas depende del buen entendimiento de la teoría básica en que se basan tales métodos.
3. Hay muchos problemas que no pueden resolverse con programas "en caja negra". Si usted es conocedor de los métodos numéricos y es hábil en la programación de computadoras, entonces tiene la capacidad de diseñar sus propios programas para resolver los problemas, sin tener que comprar un software costoso.
4. Los métodos numéricos son un vehículo eficiente para aprender a servirse de las computadoras. Es bien sabido que una forma efectiva de aprender programación consiste en escribir programas de computadora. Debido a que la mayoría de los métodos numéricos están diseñados para implementarlos en las computadoras, son ideales para tal propósito. Además, son especialmente adecuados para ilustrar el poder y las limitaciones de las computadoras. Cuando usted desarrolle en forma

satisfactoria los métodos numéricos en computadora y los aplique para resolver los problemas que de otra forma resultarían inaccesibles, usted dispondrá de una excelente demostración de cómo las computadoras sirven para su desarrollo profesional. Al mismo tiempo, aprenderá a reconocer y controlar los errores de aproximación que son inseparables de los cálculos numéricos a gran escala.

5. Los métodos numéricos son un medio para reforzar la comprensión de las matemáticas, ya que una de sus funciones es convertir las matemáticas superiores en operaciones aritméticas básicas, de esta forma se puede profundizar en los temas que de otro modo resultarían oscuros. Esta perspectiva dará como resultado un aumento de su capacidad de comprensión y entendimiento en la materia.

PT1.2 ANTECEDENTES MATEMÁTICOS

Cada parte de este libro requiere de algunos conocimientos matemáticos, por lo que el material introductorio de cada parte comprende una sección que incluye los fundamentos matemáticos. Como la parte uno, que está dedicada a aspectos básicos sobre las matemáticas y la computación, en esta sección no se revisará ningún tema matemático específico. En vez de ello se presentan los temas del contenido matemático que se cubren en este libro. Éstos se resumen en la figura PT1.2 y son:

1. *Raíces de ecuaciones* (figura PT1.2*a*). Estos problemas se relacionan con el valor de una variable o de un parámetro que satisface una ecuación no lineal. Son especialmente valiosos en proyectos de ingeniería, donde con frecuencia es imposible resolver explícitamente las ecuaciones para los parámetros definidos.
2. *Sistemas de ecuaciones algebraicas lineales* (figura PT1.2*b*). En esencia, se trata de problemas similares a los de raíces de ecuaciones, en el sentido de que están relacionados con valores que satisfacen las ecuaciones. Sin embargo, en lugar de satisfacer una sola ecuación, se busca un conjunto de valores que satisfaga simultáneamente un conjunto de ecuaciones algebraicas lineales, las cuales surgen en el contexto de una gran variedad de problemas y en todas las disciplinas de la ingeniería. En particular, se originan a partir de modelos matemáticos de grandes sistemas de elementos interrelacionados, tal como estructuras, circuitos eléctricos y redes de flujo; aunque también se llegan a encontrar en otras áreas de los métodos numéricos como el ajuste de curvas y las ecuaciones diferenciales.
3. *Optimización* (figura PT1.2*c*). En estos problemas se trata de determinar el valor o los valores de una variable independiente que corresponden al “mejor” o al valor óptimo de una función. De manera que, como se observa en la figura PT1.2*c*, la optimización considera la identificación de máximos y mínimos. Tales problemas se presentan comúnmente en el contexto del diseño en ingeniería. También surgen en otros métodos numéricos. Nosotros nos ocuparemos de la optimización tanto para una sola variable sin restricciones como para varias variables sin restricciones. También describiremos la optimización restringida dando especial énfasis a la programación lineal.
4. *Ajuste de curvas* (figura PT1.2*d*). A menudo se tendrá que ajustar curvas a un conjunto de datos representados por puntos. Las técnicas desarrolladas para tal propósito se dividen en dos categorías generales: regresión e interpolación. La primera se emplea cuando hay un significativo grado de error asociado con los datos; con frecuencia los datos experimentales son de este tipo. Para estas situaciones, la estrategia es encontrar una curva que represente la tendencia general de los datos, sin necesidad de tocar los puntos individuales. En contraste, la interpolación se utiliza cuando el objetivo es determinar valores intermedios entre datos que estén, relativamente, libres de error. Tal es el caso de la información tabulada. En dichas situaciones, la estrategia consiste en ajustar una curva directamente mediante los puntos obtenidos como datos y usar la curva para predecir valores intermedios.
5. *Integración* (figura PT1.2*e*). Como hemos representado gráficamente, la interpretación de la integración numérica es la determinación del área bajo la curva. La integración tiene diversas aplicaciones en la práctica de la ingeniería, que van desde la determinación de los centroides de objetos con formas extrañas, hasta el cálculo de cantidades totales basadas en conjuntos de medidas discretas. Además, las fórmulas de integración numérica desempeñan un papel importante en la solución de ecuaciones diferenciales.

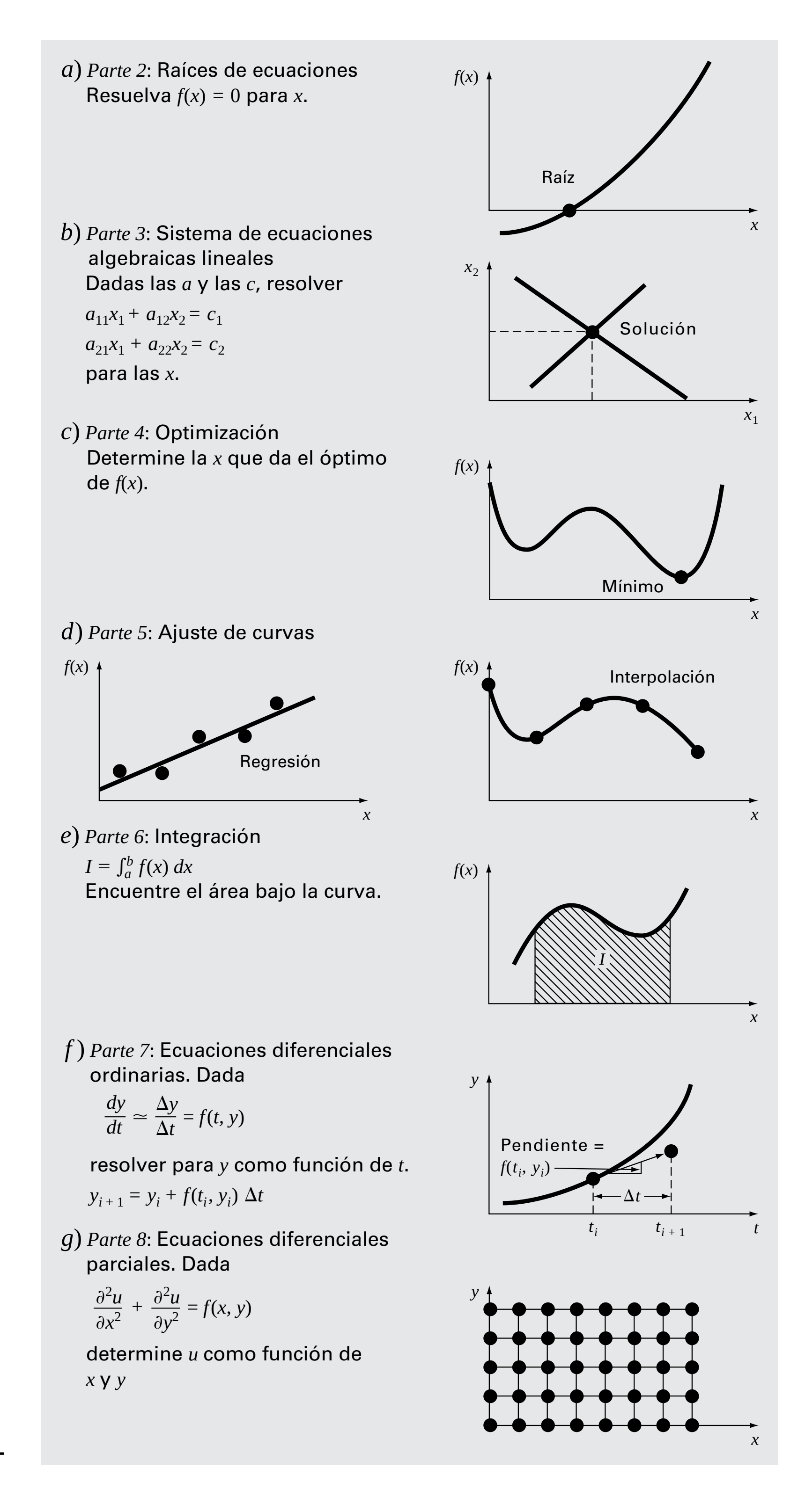


Figura PT1.2
Resumen de los métodos numéricos que se consideran en este libro.

6. *Ecuaciones diferenciales ordinarias* (figura PT1.2*f*). Éstas tienen una enorme importancia en la práctica de la ingeniería, lo cual se debe a que muchas leyes físicas están expresadas en términos de la razón de cambio de una cantidad, más que en términos de la cantidad misma. Entre los ejemplos tenemos desde los modelos de predicción demográfica (razón de cambio de la población), hasta la aceleración de un cuerpo que cae (razón de cambio de la velocidad).

Se tratan dos tipos de problemas: problemas con valor inicial y problemas con valores en la frontera. Además veremos el cálculo de valores propios.

7. *Ecuaciones diferenciales parciales* (figura PT1.2*g*). Las ecuaciones diferenciales parciales son usadas para caracterizar sistemas de ingeniería, en los que el comportamiento de una cantidad física se expresa en términos de su razón de cambio con respecto a dos o más variables independientes. Entre los ejemplos tenemos la distribución de temperatura en estado estacionario sobre una placa caliente (espacio bidimensional) o la temperatura variable con el tiempo de una barra caliente (tiempo y una dimensión espacial). Para resolver numéricamente las ecuaciones diferenciales parciales se emplean dos métodos bastante diferentes. En el presente texto haremos énfasis en los métodos de las diferencias finitas que aproximan la solución usando puntos discretos (figura PT1.2*g*). No obstante, también presentaremos una introducción a los métodos de elementos finitos, los cuales usan una aproximación con piezas discretas.

PT1.3 ORIENTACIÓN

Resulta útil esta orientación antes de proceder a la introducción de los métodos numéricos. Lo que sigue está pensado como una vista general del material contenido en la parte uno. Se incluyen, además, algunos objetivos como ayuda para concentrar el esfuerzo del lector en el estudio de los temas.

PT1.3.1 Alcance y presentación preliminar

La figura PT1.3 es una representación esquemática del material contenido en la parte uno. Este diagrama se elaboró para ofrecer un panorama global de esta parte del libro. Se considera que un sentido de “imagen global” resulta importante para desarrollar una verdadera comprensión de los métodos numéricos. Al leer un texto es posible que se pierda uno en los detalles técnicos. Siempre que el lector perciba que está perdiendo la “imagen global” vuelva a la figura PT1.3 para orientarse nuevamente. Cada parte de este libro contiene una figura similar.

La figura PT1.3 también sirve como una breve revisión inicial del material que se cubre en la parte uno. El *capítulo 1* está diseñado para orientarle en los métodos numéricos y para motivarlo mostrándole cómo se utilizan dichas técnicas, en el proceso de elaborar modelos matemáticos aplicados a la ingeniería. El *capítulo 2* es una introducción y un repaso de los aspectos de computación que están relacionados con los métodos numéricos y presenta las habilidades de programación que se deben adquirir para explotar de manera eficiente la siguiente información. Los *capítulos 3* y *4* se ocupan del importante tema del análisis del error, que debe entenderse bien para el uso efectivo de los métodos numéricos. Además, se incluye un *epílogo* que presenta los elementos de juicio que tienen una gran importancia para el uso efectivo de los métodos numéricos.

PT1.3.2 Metas y objetivos

Objetivos de estudio Al terminar la parte uno el lector deberá estar preparado para aventurarse en los métodos numéricos. En general, habrá adquirido una comprensión fundamental de la importancia de las computadoras y del papel que desempeñan las aproximaciones y los errores en el uso y desarrollo de los métodos numéricos. Además de estas metas generales, deberá dominar cada uno de los objetivos de estudio específicos que se muestran en la tabla PT1.1.

Objetivos de cómputo Al terminar de estudiar la parte uno, usted deberá tener suficientes habilidades en computación para desarrollar su propio software para los métodos numéricos de este texto. También será capaz de desarrollar programas de computadora bien estructurados y confiables basándose en pseudocódigos, diagramas de flujo u otras formas de algoritmo. Usted deberá desarrollar la capacidad de documentar sus programas de manera que sean utilizados en forma eficiente por otros usuarios. Por último, además de sus propios programas, usted deberá usar paquetes de software junto con este libro. Programas como Excel, Mathcad o MATLAB®, de The MathWorks, Inc., son ejemplos de dicho software. Usted deberá estar familiarizado con ellos, ya que será más cómodo utilizarlos para resolver después los problemas numéricos de este texto.

Tabla PT1.1 Objetivos específicos de estudio de la parte uno.

1. Reconocer la diferencia entre soluciones analíticas y numéricas.
2. Entender cómo las leyes de conservación se emplean para desarrollar modelos matemáticos de sistemas físicos.
3. Definir diseño modular y *top-down*.
4. Definir las reglas para la programación estructurada.
5. Ser capaz de elaborar programas estructurados y modulares en un lenguaje de alto nivel.
6. Saber cómo se traducen los diagramas de flujo estructurado y el pseudocódigo al código de un lenguaje de alto nivel.
7. Empezar a familiarizarse con cualquier software que usará junto con este texto.
8. Reconocer la diferencia entre error de truncamiento y error de redondeo.
9. Comprender los conceptos de cifras significativas, exactitud y precisión.
10. Conocer la diferencia entre error relativo verdadero ε_v, error relativo aproximado ε_a y error aceptable ε_s y entender cómo ε_a y ε_s sirven para terminar un proceso iterativo.
11. Entender cómo se representan los números en las computadoras y cómo tal representación induce errores de redondeo. En particular, conocer la diferencia entre precisión simple y extendida.
12. Reconocer cómo la aritmética de la computadora llega a presentar y amplificar el error de redondeo en los cálculos. En particular, apreciar el problema de la cancelación por sustracción.
13. Saber cómo la serie de Taylor y su residuo se emplean para representar funciones continuas.
14. Conocer la relación entre diferencias finitas divididas y derivadas.
15. Ser capaz de analizar cómo los errores se propagan a través de las relaciones funcionales.
16. Estar familiarizado con los conceptos de estabilidad y condición.
17. Familiarizarse con las consideraciones que se describen en el epílogo de la parte uno.

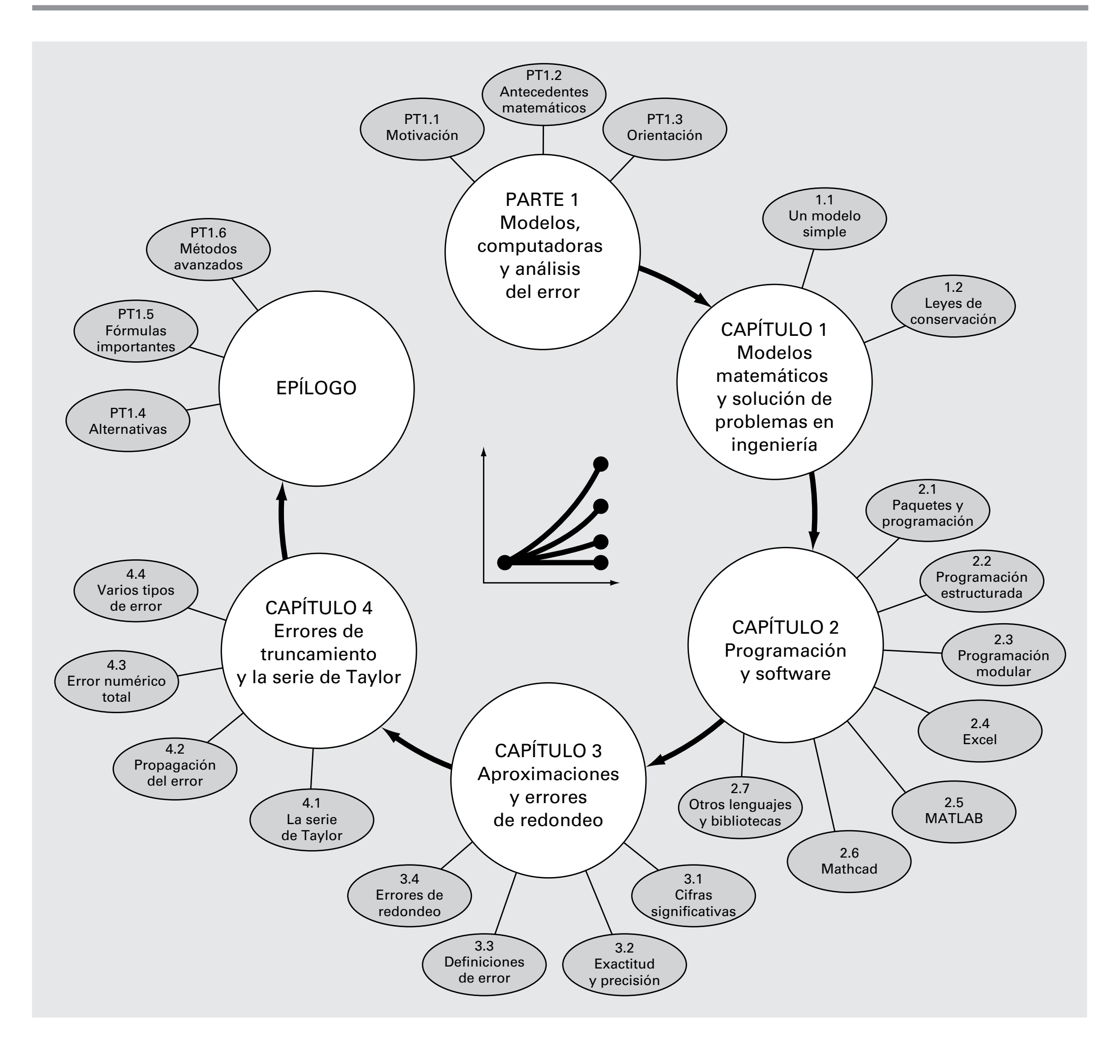


Figura PT1.3
Esquema de la organización del material en la parte uno: Modelos, computadoras y análisis del error.

CAPÍTULO 1

Modelos matemáticos y solución de problemas en ingeniería

El conocimiento y la comprensión son prerrequisitos para la aplicación eficaz de cualquier herramienta. Por ejemplo, no importa lo impresionante que sea la caja de herramientas, será muy díficil reparar un automóvil si no se entiende cómo funciona.

Ésta es una realidad, particularmente cuando se utilizan computadoras para resolver problemas de ingeniería. Aunque las computadoras tienen una gran utilidad, son prácticamente inútiles si no se comprende el funcionamiento de los sistemas de ingeniería.

Esta comprensión inicialmente es empírica —es decir, se adquiere por observación y experimentación—. Sin embargo, aunque esta información obtenida de manera empírica resulta esencial, sólo estamos a la mitad del camino. Durante muchos años de observación y experimentación, los ingenieros y los científicos han advertido que ciertos aspectos de sus estudios empíricos ocurren una y otra vez. Este comportamiento general puede expresarse como las leyes fundamentales que engloba, en esencia, el conocimiento acumulado de la experiencia pasada. Así, muchos problemas de ingeniería se resuelven con el empleo de un doble enfoque: el empirismo y el análisis teórico (figura 1.1).

Debe destacarse que ambos están estrechamente relacionados. Conforme se obtienen nuevas mediciones, las generalizaciones llegan a modificarse o aun a descubrirse otras nuevas. De igual manera, las generalizaciones tienen una gran influencia en la experimentación y en las observaciones. En lo particular, las generalizaciones sirven para organizar principios que se utilizan para sintetizar los resultados de observaciones y experimentos en un sistema coherente y comprensible, del que se pueden obtener conclusiones. Desde la perspectiva de la solución de un problema de ingeniería, el sistema es aún más útil cuando el problema se expresa por medio de un modelo matemático.

El primer objetivo de este capítulo consiste en introducir al lector a la modelación matemática y su papel en la solución de problemas en ingeniería. Se mostrará también la forma en que los métodos numéricos figuran en el proceso.

1.1 UN MODELO MATEMÁTICO SIMPLE

Un *modelo matemático* se define, de manera general, como una formulación o una ecuación que expresa las características esenciales de un sistema físico o de un proceso en términos matemáticos. En general, el modelo se representa mediante una relación funcional de la forma:

$$\begin{matrix}\text{Variable}\\ \text{dependiente}\end{matrix} = f\left(\begin{matrix}\text{variables}\\ \text{independientes}\end{matrix}, \text{parámetros}, \begin{matrix}\text{funciones}\\ \text{de fuerza}\end{matrix}\right) \tag{1.1}$$

donde la *variable dependiente* es una característica que generalmente refleja el comportamiento o estado de un sistema; las *variables independientes* son, por lo común, dimensiones como tiempo y espacio, a través de las cuales se determina el comportamiento del sistema; los *parámetros* son el reflejo de las propiedades o la composición del sistema, y las *funciones de fuerza* son influencias externas que actúan sobre el sistema.

La expresión matemática real de la ecuación (1.1) va desde una simple relación algebraica hasta un enorme y complicado grupo de ecuaciones diferenciales. Por ejemplo, a través de sus observaciones, Newton formuló su segunda ley del movimiento, la cual establece que la razón de cambio del movimiento de un cuerpo, es igual a la fuerza resultante que actúa sobre él. La expresión matemática, o el modelo, de la segunda ley es la ya conocida ecuación

$$F = ma \tag{1.2}$$

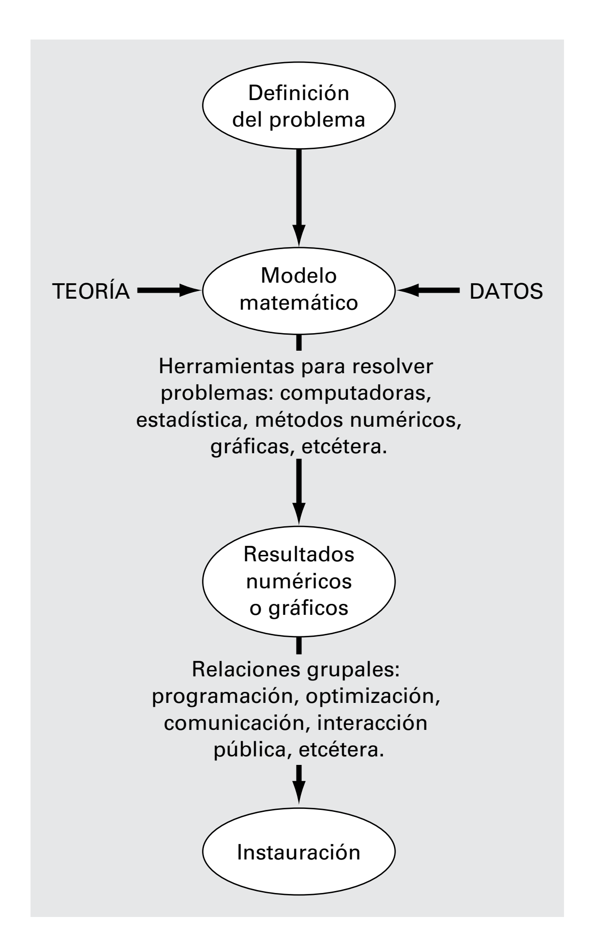


Figura 1.1
Proceso de solución de problemas en ingeniería.

donde F es la fuerza neta que actúa sobre el objeto (N o kg m/s^2), m es la masa del objeto (kg) y a es su aceleración (m/s^2).

La segunda ley puede escribirse en el formato de la ecuación (1.1), dividiendo, simplemente, ambos lados entre m para obtener

$$a = \frac{F}{m} \tag{1.3}$$

donde a es la variable dependiente que refleja el comportamiento del sistema, F es la función de fuerza y m es un parámetro que representa una propiedad del sistema. Observe que en este caso específico no existe variable independiente porque aún no se predice cómo varía la aceleración con respecto al tiempo o al espacio.

La ecuación (1.3) posee varias de las características típicas de los modelos matemáticos del mundo físico:

1. Describe un proceso o sistema natural en términos matemáticos.
2. Representa una idealización y una simplificación de la realidad. Es decir, ignora los detalles insignificantes del proceso natural y se concentra en sus manifestaciones esenciales. Por ende, la segunda ley de Newton no incluye los efectos de la relatividad, que tienen una importancia mínima cuando se aplican a objetos y fuerzas que interactúan sobre o alrededor de la superficie de la Tierra, a velocidades y en escalas visibles a los seres humanos.
3. Finalmente, conduce a resultados reproducibles y, en consecuencia, llega a emplearse con la finalidad de predecir. Por ejemplo, dada la fuerza aplicada sobre un objeto de masa conocida, la ecuación (1.3) se emplea para calcular la aceleración.

Debido a su forma algebraica sencilla, la solución de la ecuación (1.2) se obtiene con facilidad. Sin embargo, es posible que otros modelos matemáticos de fenómenos físicos sean mucho más complejos puedan ser resueltos con exactitud, o que requieran para su solución de técnicas matemáticas más sofisticadas que la simple álgebra. Para ilustrar un modelo más complicado de este tipo, se utiliza la segunda ley de Newton para determinar la velocidad final de la caída libre de un cuerpo que se encuentra cerca de la superficie de la Tierra. El cuerpo en caída libre será el de un paracaidista (figura 1.2). Un modelo para este caso se obtiene expresando la aceleración como la razón de cambio de la velocidad con respecto al tiempo (dv/dt), y sustituyendo en la ecuación (1.3). Se tiene

$$\frac{dv}{dt} = \frac{F}{m} \tag{1.4}$$

donde v es la velocidad (m/s) y t es el tiempo (s). Así, la masa multiplicada por la razón de cambio de la velocidad es igual a la fuerza neta que actúa sobre el cuerpo. Si la fuerza neta es positiva, el cuerpo se acelerará. Si es negativa, el cuerpo se desacelerará. Si la fuerza neta es igual a cero, la velocidad del cuerpo permanecerá constante.

Ahora expresemos la fuerza neta en términos de variables y parámetros mensurables. Para un cuerpo que cae a distancias cercanas a la Tierra (figura 1.2), la fuerza neta está compuesta por dos fuerzas contrarias: la atracción hacia abajo debida a la gravedad F_D y la fuerza hacia arriba debida a la resistencia del aire F_U:

$$F = F_D + F_U \tag{1.5}$$

Si a la fuerza hacia abajo se le asigna un signo positivo, la segunda ley de Newton puede ser utilizada para expresar la fuerza debida a la gravedad como

$$F_D = mg \tag{1.6}$$

donde g es la constante gravitacional, o la aceleración debida a la gravedad, que es aproximadamente igual a 9.81 m/s^2.

La resistencia del aire puede expresarse de varias maneras. Una forma sencilla consiste en suponer que es linealmente proporcional a la velocidad,[1] y que actúa en dirección hacia arriba tal como

$$F_U = -cv \tag{1.7}$$

donde c es una constante de proporcionalidad llamada *coeficiente de arrastre* (o *resistencia*) (kg/s). Así, cuanto mayor sea la velocidad de caída, mayor será la fuerza hacia arriba debida a la resistencia del aire. El parámetro c toma en cuenta las propiedades del objeto que cae, como su forma o la aspereza de su superficie, que afectan la resistencia del aire. En este caso, c podría ser función del tipo de traje o de la orientación usada por el paracaidista durante la caída libre.

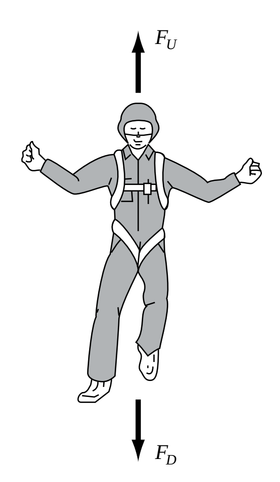


Figura 1.2 Representación esquemática de las fuerzas que actúan sobre un paracaidista en descenso. F_D es la fuerza hacia abajo debida a la atracción de la gravedad. F_U es la fuerza hacia arriba debida a la resistencia del aire.

La fuerza neta es la diferencia entre las fuerzas hacia abajo y las fuerzas hacia arriba. Por lo tanto, combinando las ecuaciones (1.4) a (1.7), se obtiene

$$\frac{dv}{dt} = \frac{mg - cv}{m} \tag{1.8}$$

o simplificando el lado derecho de la igualdad,

$$\frac{dv}{dt} = g - \frac{c}{m}v \tag{1.9}$$

La ecuación (1.9) es un modelo que relaciona la aceleración de un cuerpo que cae con las fuerzas que actúan sobre él. Se trata de una *ecuación diferencial* porque está escrita en términos de la razón de cambio diferencial (dv/dt) de la variable que nos interesa predecir. Sin embargo, en contraste con la solución de la segunda ley de Newton en la ecuación (1.3), la solución exacta de la ecuación (1.9) para la velocidad del paracaidista que cae no puede obtenerse mediante simples manipulaciones algebraicas. Es necesario emplear técnicas más avanzadas, del cálculo, para obtener una solución exacta o analítica. Por ejemplo, si inicialmente el paracaidista está en reposo ($v = 0$ en $t = 0$), se utiliza el cálculo integral para resolver la ecuación (1.9), así

$$v(t) = \frac{gm}{c}\left(1 - e^{-(c/m)t}\right) \tag{1.10}$$

Note que la ecuación (1.10) es un ejemplo de la forma general de la ecuación (1.1), donde $v(t)$ es la variable dependiente, t es la variable independiente, c y m son parámetros, y g es la función de fuerza.

EJEMPLO 1.1 Solución analítica del problema del paracaidista que cae

Planteamiento del problema Un paracaidista con una masa de 68.1 kg salta de un globo aerostático fijo. Aplique la ecuación (1.10) para calcular la velocidad antes de que se abra el paracaídas. Considere que el coeficiente de arrastre es igual a 12.5 kg/s.

Solución Al sustituir los valores de los parámetros en la ecuación (1.10) se obtiene

$$v(t) = \frac{9.81(68.1)}{12.5}\left(1 - e^{-(12.5/68.1)t}\right) = 53.44\left(1 - e^{-0.18355t}\right)$$

[1] De hecho, la relación es realmente no lineal y podría ser representada mejor por una relación con potencias como $F_U = -cv^2$. Al final de este capítulo investigaremos, en un ejercicio, de qué manera influyen estas no linealidades en el modelo.

que sirve para calcular la velocidad del paracaidista a diferentes tiempos, tabulando se tiene el cuadro de la derecha.

De acuerdo con el modelo, el paracaidista acelera rápidamente (figura 1.3). Se alcanza una velocidad de 44.92 m/s después de 10 s. Observe también que, después de un tiempo suficientemente grande, alcanza una velocidad constante llamada *velocidad terminal* o *velocidad límite* de 53.44 m/s. Esta velocidad es constante porque después de un tiempo la fuerza de gravedad estará en equilibrio con la resistencia del aire. Entonces, la fuerza neta es cero y cesa la aceleración.

t, s	*v*, m/s
0	0.00
2	16.42
4	27.80
6	35.68
8	41.14
10	44.92
12	47.54
∞	53.44

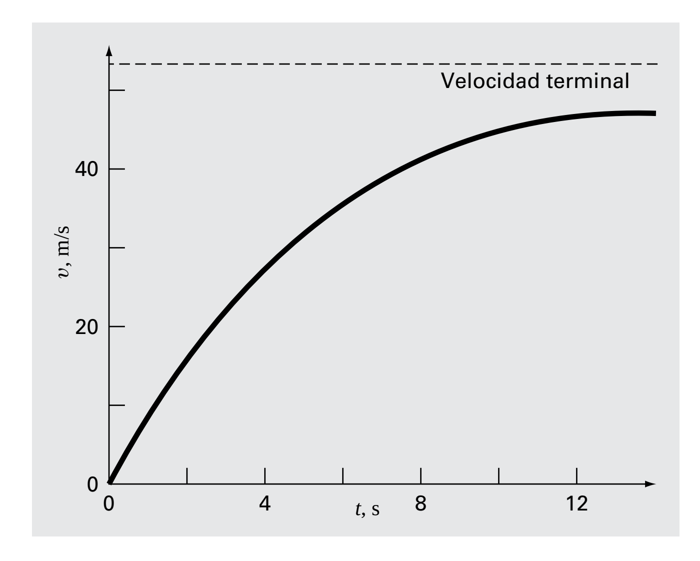


Figura 1.3
Solución analítica al problema del paracaidista que cae según se calcula en el ejemplo 1.1. La velocidad aumenta con el tiempo y tiende asintóticamente a una velocidad terminal.

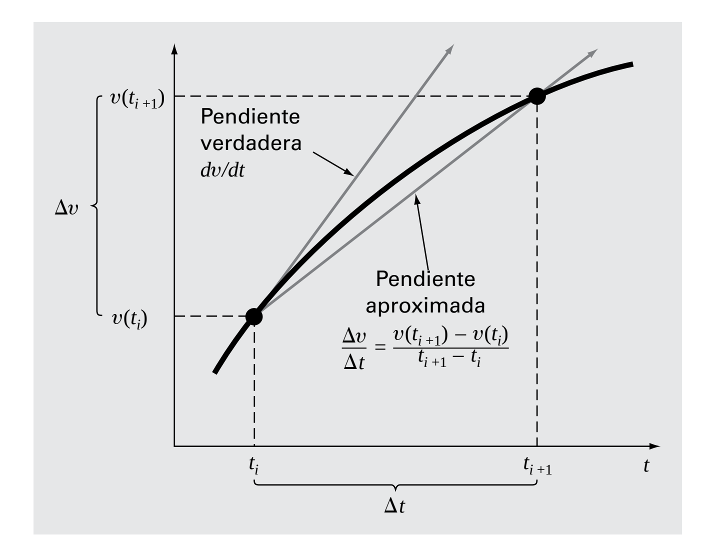


Figura 1.4
Uso de una diferencia finita para aproximar la primera derivada de v con respecto a t.

A la ecuación (1.10) se le llama *solución analítica* o *exacta* ya que satisface con exactitud la ecuación diferencial original. Por desgracia, hay muchos modelos matemáticos que no pueden resolverse con exactitud. En muchos de estos casos, la única alternativa consiste en desarrollar una solución numérica que se aproxime a la solución exacta.

Como ya se mencionó, los *métodos numéricos* son aquellos en los que se reformula el problema matemático para lograr resolverlo mediante operaciones aritméticas. Esto puede ilustrarse para el caso de la segunda ley de Newton, observando que a la razón de cambio de la velocidad con respecto al tiempo se puede aproximar mediante (figura 1.4):

$$\frac{dv}{dt} \cong \frac{\Delta v}{\Delta t} = \frac{v(t_{i+1}) - v(t_i)}{t_{i+1} - t_i} \tag{1.11}$$

donde Δv y Δt son diferencias en la velocidad y en el tiempo, respectivamente, calculadas sobre intervalos finitos, $v(t_i)$ es la velocidad en el tiempo inicial t_i, y $v(t_{i+1})$ es la velocidad algún tiempo más tarde t_{i+1}. Observe que $dv/dt \cong \Delta v/\Delta t$ es aproximado porque Δt es finito. Al recordar los cursos de cálculo tenemos que

$$\frac{dv}{dt} = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t}$$

La ecuación (1.11) representa el proceso inverso.

A la ecuación (1.11) se le denomina una aproximación en *diferencia finita dividida* de la derivada en el tiempo t_i. Al sustituir en la ecuación (1.9), tenemos

$$\frac{v(t_{i+1}) - v(t_i)}{t_{i+1} - t_i} = g - \frac{c}{m} v(t_i)$$

Esta ecuación se reordena para obtener

$$v(t_{i+1}) = v(t_i) + \left[g - \frac{c}{m} v(t_i) \right] (t_{i+1} - t_i) \tag{1.12}$$

Note que el término entre corchetes es el lado derecho de la propia ecuación diferencial [ecuación (1.9)]. Es decir, este término nos da un medio para calcular la razón de cambio o la pendiente de v. Así, la ecuación diferencial se ha transformado en una ecuación que puede utilizarse para determinar algebraicamente la velocidad en t_{i+1}, usando la pendiente y los

valores anteriores de v y t. Si se da un valor inicial para la velocidad en algún tiempo t_i, es posible calcular con facilidad la velocidad en un tiempo posterior t_{i+1}. Este nuevo valor de la velocidad en t_{i+1} sirve para calcular la velocidad en t_{i+2} y así sucesivamente. Es decir, a cualquier tiempo,

$$\begin{matrix}\text{valor}\\ \text{nuevo}\end{matrix} = \text{valor anterior} + \text{pendiente} \times \begin{matrix}\text{tamaño}\\ \text{del paso}\end{matrix}$$

Observe que esta aproximación formalmente se conoce como *método de Euler.*

EJEMPLO 1.2 Solución numérica al problema de la caída de un paracaidista

Planteamiento del problema Realice el mismo cálculo que en el ejemplo 1.1, pero usando la ecuación (1.12) para obtener la velocidad. Emplee un tamaño de paso de 2 s para el cálculo.

Solución Al empezar con los cálculos ($t_i = 0$), la velocidad del paracaidista es igual a cero. Con esta información y los valores de los parámetros del ejemplo 1.1, se utiliza la ecuación (1.12) para calcular la velocidad en $t_{i+1} = 2$ s:

$$v = 0 + \left[9.81 - \frac{12.5}{68.1}(0)\right]2 = 19.62 \text{ m/s}$$

Para el siguiente intervalo (de $t = 2$ a 4 s) se repite el cálculo y se obtiene

$$v = 19.62 + \left[9.81 - \frac{12.5}{68.1}(19.62)\right]2 = 32.04 \text{ m/s}$$

t, s	*v*, m/s
0	0.00
2	19.62
4	32.04
6	39.90
8	44.87
10	48.02
12	50.01
∞	53.44

Se continúa con los cálculos de manera similar para obtener los valores de la tabla de la izquierda.

Los resultados se muestran gráficamente en la figura 1.5, junto con la solución exacta. Como se puede ver, el método numérico se aproxima bastante a la solución exacta. Sin embargo, debido a que se emplean segmentos de rectas para aproximar una función que es una curva continua, hay algunas diferencias entre los dos resultados. Una forma de reducir estas diferencias consiste en usar un tamaño de paso menor. Por ejemplo, si se aplica la ecuación (1.12) con intervalos de 1 s, se obtendría un error menor, ya que los segmentos de recta estarían un poco más cerca de la verdadera solución. Con los cálculos manuales, el esfuerzo asociado al usar incrementos cada vez más pequeños haría poco prácticas tales soluciones numéricas. No obstante, con la ayuda de una computadora personal es posible efectuar fácilmente un gran número de cálculos; por lo tanto, se puede modelar con más exactitud la velocidad del paracaidista que cae, sin tener que resolver exactamente la ecuación diferencial.

Como se vio en el ejemplo anterior, obtener un resultado numérico más preciso tiene un costo en términos del número de cálculos. Cada división a la mitad del tamaño de paso para lograr mayor precisión nos lleva a duplicar el número de cálculos. Como vemos, existe un costo inevitable entre la exactitud y la cantidad de operaciones. Esta relación es de gran importancia en los métodos numéricos y constituyen un tema relevante de este libro. En consecuencia, hemos dedicado el epílogo de la parte uno para ofrecer una introducción a dicho tipo de relaciones.

1.2 LEYES DE CONSERVACIÓN E INGENIERÍA

Aparte de la segunda ley de Newton, existen otros principios importantes en ingeniería. Entre los más importantes están las leyes de conservación. Aunque éstas son fundamentales en una gran variedad de complicados y poderosos modelos matemáticos, las leyes de conservación en la ciencia y en la ingeniería conceptualmente son fáciles de entender. Todas ellas se pueden reducir a

$$\text{Cambio} = \text{incremento} - \text{decremento} \tag{1.13}$$

Éste es precisamente el formato que empleamos al usar la segunda ley de Newton para desarrollar un equilibrio de fuerzas en la caída del paracaidista [ecuación (1.8)].

Pese a su sencillez, la ecuación (1.13) representa una de las maneras fundamentales en que las leyes de conservación se emplean en ingeniería —esto es, predecir cambios con respecto al tiempo—. Nosotros le daremos a la ecuación (1.13) el nombre especial de cálculo de *variable-tiempo* (o *transitorio*).

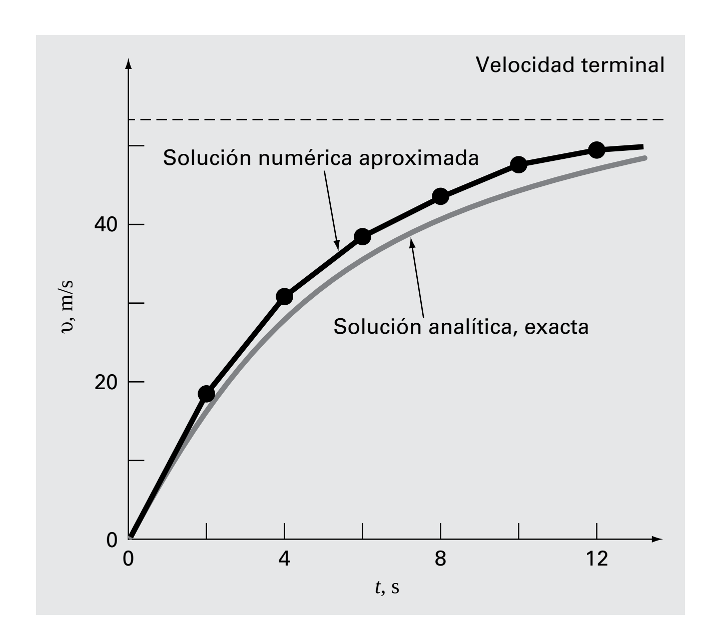


Figura 1.5
Comparación de las soluciones numéricas y analíticas para el problema del paracaidista que cae.

Además de la predicción de cambios, las leyes de conservación se aplican también en casos en los que no existe cambio. Si el cambio es cero, la ecuación (1.3) será

$$\text{Cambio} = 0 = \text{incremento} - \text{decremento}$$

o bien,

$$\text{Incremento} = \text{decremento} \tag{1.14}$$

Así, si no ocurre cambio alguno, el incremento y el decremento deberán estar en equilibrio. Este caso, al que también se le da una denominación especial —cálculo en *estado estacionario*—, tiene diversas aplicaciones en ingeniería. Por ejemplo, para el flujo de un fluido incompresible en estado estacionario a través de tuberías, el flujo de entrada debe estar en equilibrio con el flujo de salida, esto es

$$\text{Flujo de entrada} = \text{flujo de salida}$$

Para la unión de tuberías de la figura 1.6, esta ecuación de equilibrio se utiliza para calcular el flujo de salida de la cuarta tubería, que debe ser de 60.

Para la caída del paracaidista, las condiciones del estado estacionario deberían corresponder al caso en que la fuerza neta fuera igual a cero o [ecuación (1.8) con $dv/dt = 0$]

$$mg = cv \tag{1.15}$$

Así, en el estado estacionario, las fuerzas hacia abajo y hacia arriba están equilibradas, y en la ecuación (1.15) puede encontrarse la velocidad terminal

$$v = \frac{mg}{c}$$

Aunque las ecuaciones (1.13) y (1.14) pueden parecer triviales, éstas determinan las dos maneras fundamentales en que las leyes de la conservación se emplean en ingeniería. Como tales, en los capítulos siguientes serán parte importante de nuestros esfuerzos por mostrar la relación entre los métodos numéricos y la ingeniería. Nuestro primer medio para establecer tal relación son las aplicaciones a la ingeniería que aparecen al final de cada parte del libro.

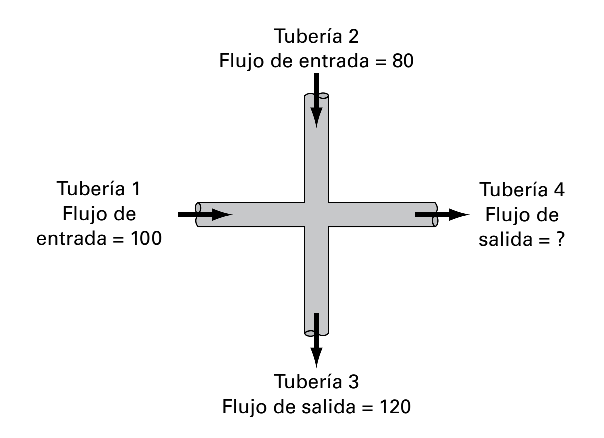


Figura 1.6
Equilibrio del flujo de un fluido incompresible en estado estacionario a través de tuberías.

En la tabla 1.1 se resumen algunos de los modelos sencillos de ingeniería y las leyes de conservación correspondientes, que constituirán la base de muchas de las aplicaciones a la ingeniería. La mayoría de aplicaciones de ingeniería química harán énfasis en el balance de masa para el estudio de los reactores. El balance de masa es una consecuencia de la conservación de la masa. Éste especifica que, el cambio de masa de un compuesto químico en un reactor, depende de la cantidad de masa que entra menos la cantidad de masa que sale.

Tabla 1.1 Dispositivos y tipos de balances que se usan comúnmente en las cuatro grandes áreas de la ingeniería. En cada caso se especifica la ley de conservación en que se fundamenta el balance.

Campo	Dispositivo	Principio aplicado	Expresión matemática
Ingeniería química	Reactores	Conservación de la masa	Balance de masa: Entrada → Salida En un periodo Δmasa = entradas – salidas
Ingeniería civil	Estructura	Conservación del *momentum*	Equilibrio de fuerzas: $+F_V$, $-F_H$, $+F_H$, $-F_V$ En cada nodo Σ fuerzas horizontales $(F_H) = 0$ Σ fuerzas verticales $(F_V) = 0$
Ingeniería mecánica	Máquina	Conservación del *momentum*	Equilibrio de fuerzas: Fuerza hacia arriba, $x = 0$, Fuerza hacia abajo $m\,\dfrac{d^2x}{dt^2}$ = Fuerza hacia abajo – fuerza hacia arriba
Ingeniería eléctrica	Circuito (+, –)	Conservación de la carga	Balance de corriente: $+i_1$, $-i_3$, $+i_2$ En cada nodo Σ corriente $(i) = 0$
		Conservación de la energía	Balance de voltaje: i_1R_1, i_2R_2, ξ, i_3R_3 Alrededor de cada malla Σ fems – Σ caída de potencial en los resistores = 0 $\Sigma\,\xi - \Sigma\,iR = 0$

TABLA 1.2 Algunos aspectos prácticos que se investigarán en las aplicaciones a la ingeniería al final de cada parte del libro.

1. *No lineal contra lineal*. Mucho de la ingeniería clásica depende de la linealización que permite soluciones analíticas. Aunque esto es con frecuencia apropiado, puede lograrse una mejor comprensión cuando se revisan los problemas no lineales.
2. *Grandes sistemas contra pequeños*. Sin una computadora, no siempre es posible examinar sistemas en que intervienen más de tres componentes. Con las computadoras y los métodos numéricos, se pueden examinar en forma más realista sistemas multicomponentes.
3. *No ideal contra ideal*. En ingeniería abundan las leyes idealizadas. A menudo, hay alternativas no idealizadas que son más realistas pero que demandan muchos cálculos. La aproximación numérica llega a facilitar la aplicación de esas relaciones no ideales.
4. *Análisis de sensibilidad*. Debido a que están involucrados, muchos cálculos manuales requieren una gran cantidad de tiempo y esfuerzo para su correcta realización. Esto algunas veces desalienta al analista cuando realiza los múltiples cálculos que son necesarios al examinar cómo responde un sistema en diferentes condiciones. Tal análisis de sensibilidad se facilita cuando los métodos numéricos permiten que la computadora asuma la carga de cálculo.
5. *Diseño*. Determinar el comportamiento de un sistema en función de sus parámetros es a menudo una proposición sencilla. Por lo común, es más difícil resolver el problema inverso; es decir, determinar los parámetros cuando se especifica el comportamiento requerido. Entonces, los métodos numéricos y las computadoras permiten realizar esta tarea de manera eficiente.

Las aplicaciones en ingeniería civil y mecánica se enfocan al desarrollo de modelos a partir de la conservación del *momentum*. En la ingeniería civil se utilizan fuerzas en equilibrio para el análisis de estructuras como las armaduras sencillas de la tabla 1.1. El mismo principio se aplica en ingeniería mecánica, con la finalidad de analizar el movimiento transitorio hacia arriba o hacia abajo, o las vibraciones de un automóvil.

Por último, las aplicaciones en ingeniería eléctrica emplean tanto balances de corriente como de energía para modelar circuitos eléctricos. El balance de corriente, que resulta de la conservación de carga, es similar al balance del flujo representado en la figura 1.6. Así como el flujo debe equilibrarse en las uniones de tuberías, la corriente eléctrica debe estar balanceada o en equilibrio en las uniones de alambres eléctricos. El balance de energía especifica que la suma algebraica de los cambios de voltaje alrededor de cualquier malla de un circuito debe ser igual a cero. Las aplicaciones en ingeniería se proponen para ilustrar cómo se emplean actualmente los métodos numéricos en la solución de problemas en ingeniería. Estas aplicaciones nos permitirán examinar la solución a los problemas prácticos (tabla 1.2) que surgen en el mundo real. Establecer la relación entre las técnicas matemáticas como los métodos numéricos y la práctica de la ingeniería es un paso decisivo para mostrar su verdadero potencial. Examinar de manera cuidadosa las aplicaciones a la ingeniería nos ayudará a establecer esta relación.

PROBLEMAS

1.1 Utilice el cálculo para resolver la ecuación (1.9) para el caso en que la velocidad inicial, $v(0)$ es diferente de cero.

1.2 Repita el ejemplo 1.2. Calcule la velocidad en $t = 8$ s, con un tamaño de paso de *a*) 1 y *b*) 0.5 s. ¿Puede establecer algún enunciado en relación con los errores de cálculo con base en los resultados?

1.3 En vez de la relación lineal de la ecuación (1.7), elija modelar la fuerza hacia arriba sobre el paracaidista como una relación de segundo orden,

$$F_U = -c'v^2$$

donde c' = un coeficiente de arrastre de segundo orden (kg/m).

a) Con el empleo del cálculo, obtenga la solución de forma cerrada para el caso en que al inicio el saltador se encuentra en reposo ($v = 0$ en $t = 0$).

b) Repita el cálculo numérico en el ejemplo 1.2 con los mismos valores de condición inicial y de parámetros, pero con un arrastre de segundo orden. Utilice un valor de 0.22 kg/m para c'.

1.4 Para el paracaidista en caída libre con arrastre lineal, suponga un primer saltador de 70 kg con coeficiente de arrastre de 12 kg/s. Si un segundo saltador tiene un coeficiente de arrastre de 15 kg/s y una masa de 80 kg, ¿cuánto tiempo le tomará alcanzar la misma velocidad que el primero adquiera en 9 s?

1.5 Calcule la velocidad de un paracaidista en caída libre con el empleo del método de Euler para el caso en que $m = 80$ kg y $c = 10$ kg/s. Lleve a cabo el cálculo desde $t = 0$ hasta $t = 20$ s con un tamaño de paso de 1 s. Use una condición inicial en que el paracaidista tiene una velocidad hacia arriba de 20 m/s en $t = 0$. Suponga que el paracaídas se abre instantáneamente en $t = 10$ s, de modo que el coeficiente de arrastre sube a 60 kg/s.

1.6 La siguiente información está disponible para una cuenta bancaria:

Fecha	Depósitos	Retiros	Interés	Saldo
5/1				1 522.33
	220.13	327.26		
6/1				
	216.80	378.51		
7/1				
	450.35	106.80		
8/1				
	127.31	350.61		
9/1				

Observe que el dinero gana interés que se calcula como

Interés = iB_i

donde i = tasa de interés expresada como fracción mensual, y B_i es el saldo inicial al principio del mes.

a) Use la conservación de efectivo para calcular el saldo en las fechas 6/1, 7/1, 8/1 y 9/1 si la tasa de interés es 1% mensual (i = 0.01/mes). Muestre cada paso del cálculo.

b) Escriba una ecuación diferencial para el saldo de efectivo en la forma:

$$\frac{dB}{dt} = f(D(t), W(t), i)$$

donde t = tiempo (meses), $D(t)$ = depósitos como función del tiempo ($/mes), $W(t)$ = retiros como función del tiempo ($/mes). Para este caso, suponga que el interés es continuamente compuesto; es decir, interés = iB.

c) Use el método de Euler con un paso de tiempo de 0.5 meses para simular el saldo. Suponga que los depósitos y retiros se aplican uniformemente a lo largo de todo el mes.

d) Desarrolle una gráfica de saldo contra tiempo para *a*) y *c*).

1.7 La cantidad de un contaminante radiactivo distribuido uniformemente que se encuentra contenido en un reactor cerrado, se mide por su concentración c (becquerel/litro, o Bq/L). El contaminante disminuye con una tasa de decaimiento proporcional a su concentración, es decir:

tasa de decaimiento = $-kc$

donde k es una constante con unidades de día^{-1}. Entonces, de acuerdo con la ecuación (1.13), puede escribirse un balance de masa para el reactor, así:

$$\frac{dc}{dt} = -kc$$

$$\begin{pmatrix}\text{cambio}\\ \text{de la masa}\end{pmatrix} = \begin{pmatrix}\text{disminución}\\ \text{por decaimiento}\end{pmatrix}$$

a) Use el método de Euler para resolver esta ecuación desde $t = 0$ hasta 1 d, con $k = 0.175$ d^{-1}. Emplee un tamaño de paso de Δt = 0.1. La concentración en $t = 0$ es de 100 Bq/L.

b) Muestre la solución en un gráfico semilogarítmico (p. ej., ln c *versus* t) y determine la pendiente. Interprete sus resultados.

1.8 Un grupo de 35 estudiantes asiste a clase en un salón que mide 11 m por 8 m por 3 m. Cada estudiante ocupa alrededor de 0.075 m^3 y genera cerca de 80 W de calor (1 W = 1 J/s). Calcule el incremento de la temperatura del aire durante los primeros 20 minutos de la clase, si el salón está sellado y aislado por completo. Suponga que la capacidad calorífica del aire, C_v, es de 0.718 kJ/(kg K). Suponga que el aire es un gas ideal a 20 °C y 101.325 kPa. Obsérvese que el calor absorbido por el aire Q está relacionado con la masa de aire m, la capacidad calorífica, y el cambio en la temperatura, por medio de la relación siguiente:

$$Q = m\int_{T_1}^{T_2} C_v\,dT = mC_v(T_2 - T_1)$$

La masa del aire se obtiene de la ley del gas ideal:

$$PV = \frac{m}{\text{Mwt}}RT$$

donde P es la presión del gas, V es el volumen de éste, Mwt es el peso molecular del gas (para el aire, 28.97 kg/kmol) y R es la constante universal de los gases ideales [8.314 kPa m^3/(kmol K)].

1.9 Un tanque de almacenamiento contiene un líquido con profundidad y, donde $y = 0$ cuando el tanque está lleno a la mitad. El líquido se extrae con una tasa de flujo constante Q a fin de satisfacer las demandas. Se suministra el contenido a una tasa senoidal de $3Q$ sen$^2(t)$.

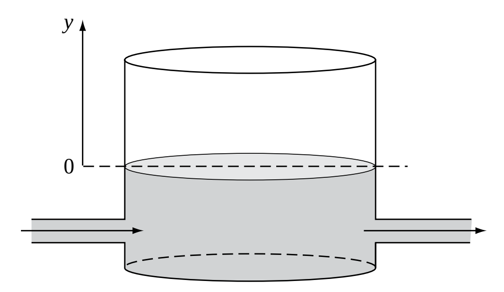


Figura P1.9

Para este sistema, la ecuación (1.13) puede escribirse como

$$\frac{d(Ay)}{dt} = 3Q\,\text{sen}^2(t) - Q$$

$$\begin{pmatrix}\text{cambio en}\\ \text{el volumen}\end{pmatrix} = (\text{flujo de entrada}) - (\text{flujo de salida})$$

o bien, como el área de la superficie A es constante

$$\frac{dy}{dt} = 3\frac{Q}{A}\,\text{sen}^2(t) - \frac{Q}{A}$$

Emplee el método de Euler para resolver cuál sería la profundidad y, desde $t = 0$ hasta 10 d, con un tamaño de paso de 0.5 d. Los valores de los parámetros son A = 1 250 m^2 y Q = 450 m^3/d. Suponga que la condición inicial es $y = 0$.

1.10 Para el mismo tanque de almacenamiento que se describe en el problema 1.9, suponga que el flujo de salida no es constante sino que la tasa depende de la profundidad. Para este caso, la ecuación diferencial para la profundidad puede escribirse como

$$\frac{dy}{dt} = 3\frac{Q}{A}\,\text{sen}^2(t) - \frac{\alpha(1+y)^{1.5}}{A}$$

Use el método de Euler para resolver cuál sería la profundidad y, desde $t = 0$ hasta 10 d, con un tamaño de paso de 0.5 d. Los valores de los parámetros son A = 1 250 m^2, Q = 450 m^3/d, y α = 150. Suponga que la condición inicial es $y = 0$.

1.11 Aplique la conservación de volumen (vea el problema 1.9) para simular el nivel de líquido en un tanque cónico de almacenamiento (figura P1.11). El líquido entra con un tasa senoidal de $Q_{ent} = 3\,\text{sen}^2(t)$, y sale de acuerdo con

$$Q_{sal} = 3(y - y_{sal})^{1.5} \qquad y > y_{sal}$$

$$Q_{sal} = 0 \qquad y \leq y_{sal}$$

donde el caudal tiene unidades de m^3/d y y = elevación de la superficie de agua sobre el fondo del tanque (m). Use el método de Euler para determinar la profundidad y desde t = 0 hasta 10 d, con un tamaño de paso de 0.5 d. Los valores de los parámetros son: r_{sup} = 2.5 m, y_{sup} = 4 m, y y_{sal} = 1 m. Suponga que el nivel está inicialmente debajo del tubo de salida con $y(0)$ = 0.8 m.

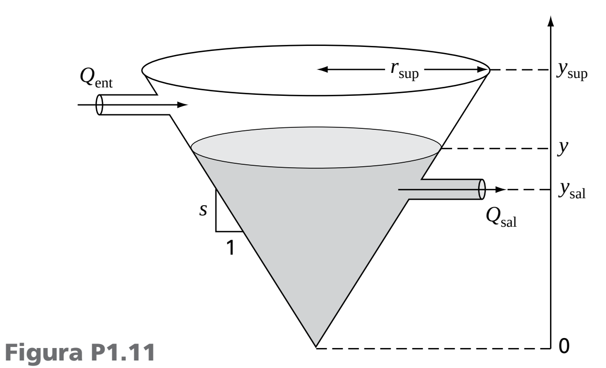


Figura P1.11

1.12 En el ejemplo del paracaidista en caída libre, se supuso que la aceleración debida a la gravedad era un valor constante. Aunque ésta es una buena aproximación cuando se estudian objetos en caída cerca de la superficie de la Tierra, la fuerza gravitacional disminuye cuando nos movemos arriba del nivel del mar. Una representación más general basada en la *ley de Newton del inverso del cuadrado* de la atracción gravitacional, se escribe como

$$g(x) = g(0)\frac{R^2}{(R+x)^2}$$

donde $g(x)$ = aceleración gravitacional a una altitud x (en m) medida hacia arriba a partir de la superficie terrestre (m/s^2), $g(0)$ = aceleración gravitacional en la superficie terrestre ($\cong$ 9.81 m/s^2) y R = el radio de la Tierra ($\cong 6.37 \times 10^6$ m).

a) En forma similar en que se obtuvo la ecuación (1.9), use un balance de fuerzas para obtener una ecuación diferencial para la velocidad como función del tiempo que utilice esta representación más completa de la gravitación. Sin embargo, para esta obtención, suponga como positiva la velocidad hacia arriba.

b) Para el caso en que el arrastre es despreciable, utilice la regla de la cadena para expresar la ecuación diferencial como función de la altitud en lugar del tiempo. Recuerde que la regla de la cadena es

$$\frac{dv}{dt} = \frac{dv}{dx}\frac{dx}{dt}$$

c) Use el cálculo para obtener la forma cerrada de la solución donde $v = v_0$ en x = 0.

d) Emplee el método de Euler para obtener la solución numérica desde x = 0 hasta 100 000 m, con el uso de un paso de 10 000 m, donde la velocidad inicial es de 1 500 m/s hacia arriba. Compare su resultado con la solución analítica.

1.13 Suponga que una gota esférica de líquido se evapora a una tasa proporcional al área de su superficie.

$$\frac{dV}{dt} = -kA$$

donde V = volumen (mm^3), t = tiempo (min), k = la tasa de evaporación (mm/min), y A = área superficial (mm^2). Emplee el método de Euler para calcular el volumen de la gota desde t = 0 hasta 10 min usando un tamaño de paso de 0.25 min. Suponga que k = 0.08 mm/min, y que al inicio la gota tiene un radio de 2.5 mm. Evalúe la validez de sus resultados por medio de determinar el radio de su volumen final calculado y la verificación de que es consistente con la tasa de evaporación.

1.14 La *ley del enfriamiento de Newton* establece que la temperatura de un cuerpo cambia con una tasa que es proporcional a la diferencia de su temperatura y la del medio que lo rodea (temperatura ambiente).

$$\frac{dT}{dt} = -k(T - T_a)$$

donde T = temperatura del cuerpo (°C), t = tiempo (min), k = constante de proporcionalidad (por minuto) y T_a = temperatura del ambiente (°C). Suponga que una tasa de café tiene originalmente una temperatura de 70 °C. Emplee el método de Euler para calcular la temperatura desde t = 0 hasta 10 min, usando un tamaño de paso de 2 min, si T_a = 20 °C y k = 0.019/min.

1.15 Como se ilustra en la figura P1.15, un *circuito RLC* consiste en tres elementos: un resistor (R), un inductor (L) y un capacitor (C). El flujo de corriente a través de cada elemento induce una caída de voltaje. La segunda ley del voltaje de Kirchhoff dice que la suma algebraica de estas caídas de voltaje alrededor de un circuito cerrado es cero,

$$iR + L\frac{di}{dt} + \frac{q}{C} = 0$$

donde i = corriente, R = resistencia, L = inductancia, t = tiempo, q = carga y C = capacitancia. Además, la corriente se relaciona con la carga como en

$$\frac{dq}{dt} = i$$

a) Si los valores iniciales son $i(0) = 0$ y $q(0)$ = 1C, use el método de Euler para resolver este par de ecuaciones diferenciales desde t = 0 hasta 0.1 s, usando un tamaño de paso de Δt = 0.01 s. Utilice los siguientes parámetros para su cálculo: R = 200 Ω, L = 5 H y $C = 10^{-4}$ F.

b) Desarrolle una gráfica de i y q contra t.

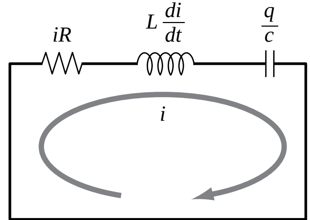


Figura P1.15

1.16 Las células cancerosas crecen en forma exponencial con un tiempo de duplicación de 20 h cuando tienen una fuente ilimitada de nutrientes. Sin embargo, conforme las células comienzan a formar un tumor de forma esférica sin abasto de sangre, el crecimiento en el centro del tumor queda limitado, y eventualmente las células empiezan a morir.

a) El crecimiento exponencial del número de células N puede expresarse como se indica, donde μ es la tasa de crecimiento de las células. Encuentre el valor de μ para las células cancerosas.

$$\frac{dN}{dt} = \mu N$$

b) Construya una ecuación que describa la tasa de cambio del volumen del tumor durante el crecimiento exponencial, dado que el diámetro de una célula individual es de 20 micras.

c) Una vez que un tipo particular de tumor excede las 500 micras de diámetro, las células del centro del tumor se mueren (pero continúan ocupando espacio en el tumor). Determine cuánto tiempo tomará que el tumor exceda ese tamaño crítico.

1.17 Se bombea un fluido por la red que se ilustra en la figura P1.17. Si $Q_2 = 0.6$, $Q_3 = 0.4$, $Q_7 = 0.2$ y $Q_8 = 0.3$ m^3/s, determine los otros flujos.

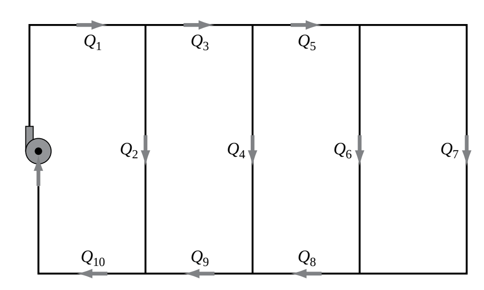


Figura P1.17

1.18 La velocidad es igual a la razón de cambio de la distancia x (m),

$$\frac{dx}{dt} = v(t) \qquad \text{(P1.18)}$$

a) Sustituya la ecuación (1.10) y desarrolle una solución analítica para distancia como función del tiempo. Suponga que $x(0) = 0$.

b) Use el método de Euler para integrar numéricamente las ecuaciones (P1.18) y (1.9) con objeto de determinar tanto la velocidad como la distancia de caída como función del tiempo para los primeros 10 s de caída libre usando los mismos parámetros que en el ejemplo 1.2.

c) Trace una gráfica de sus resultados numéricos junto con las soluciones analíticas.

1.19 Usted está trabajando como investigador de escena de crimen, y debe predecir la temperatura de una víctima de homicidio en un periodo de 5 horas. Usted sabe que el cuarto donde se encontró la víctima estaba a 10 °C cuando se descubrió el cadáver.

a) Use la *ley del enfriamiento de Newton* (problema 1.14) y el método de Euler para calcular la temperatura del cuerpo de la víctima para el periodo de 5 horas, usando valores de $k = 0.12$/h y $\Delta t = 0.5$ h. Suponga que la temperatura del cuerpo de la víctima en el momento de la muerte era de 37 °C y que la temperatura del cuarto estaba a un valor constante de 10 °C durante el periodo de 5 horas.

b) La investigación posterior revela que la temperatura del cuarto en realidad había caído linealmente de 20 a 10 °C durante el periodo de 5 horas. Repita el mismo cálculo que en el inciso *a*), pero incluya esta nueva información.

c) Compare los resultados del inciso *a*) y del inciso *b*) graficándolos en la misma gráfica.

1.20 Suponga que un paracaidista con arrastre lineal ($m = 70$ kg, $c = 12.5$ kg/s) salta de un avión que vuela a una altitud de un kilómetro con una velocidad horizontal de 180 m/s relativa a la tierra.

a) Escriba un sistema de cuatro ecuaciones diferenciales para x, y, $v_x = dx/dt$ y $v_y = dy/dt$.

b) Si la posición horizontal inicial se define como $x = 0$, use métodos de Euler con $\Delta t = 1$ s para calcular la posición del paracaidista en los primeros 10 s.

c) Desarrolle gráficas de y contra t y y contra x. Use la gráfica para estimar gráficamente cuándo y dónde chocaría con la tierra el paracaidista si el paracaídas no se abriera.

1.21 Como se observa en el problema 1.3, la fuerza de arrastre se representa con más exactitud como dependiente del cuadrado de la velocidad. Una representación más fundamental de la fuerza de arrastre, que supone condiciones turbulentas (es decir, un alto número de Reynolds) se puede formular como

$$F_d = -\frac{1}{2}\rho A C_d v|v|$$

donde F_d = la fuerza de arrastre (N), ρ = densidad del fluido (kg/m^3), A = el área frontal del objeto en un plano perpendicular a la dirección del movimiento (m^2), v = velocidad (m/s) y C_d = coeficiente adimensional de arrastre.

a) Escriba el par de ecuaciones diferenciales para la velocidad y la posición (vea el problema 1.18) para describir el movimiento vertical de una esfera con diámetro d (m) y una densidad de ρ_s (kg/m^3). La ecuación diferencial para la velocidad se debe escribir como función del diámetro de la esfera.

b) Use el método de Euler con un tamaño de paso $\Delta t = 2$ s para calcular la posición y la velocidad de una esfera durante los primeros 14 s. Utilice Los siguientes parámetros en su cálculo: d = 120 cm, $\rho = 1.3$ kg/m^3, $\rho_s = 2\,700$ kg/m^3 y $C_d = 0.47$. Suponga que la esfera tiene las siguientes condiciones iniciales: $x(0) = 100$ m y $v(0) = -40$ m/s.

c) Desarrolle una gráfica de sus resultados (es decir, y y v contra t) y úsela para estimar gráficamente cuándo chocará la esfera con el suelo.

d) Calcule el valor del coeficiente general de arrastre de segundo orden c_d'(kg/m). Observe que, como se describe en el problema 1.3, el coeficiente general de arrastre de segundo orden es el término de velocidad en la ecuación diferencial final que multiplica el término $v|v|$.

1.22 Como se ilustra en la figura P1.22, una partícula esférica que se sedimenta a través de un fluido quiescente está sujeta a tres fuerzas: la fuerza descendente de la gravedad (F_G), la fuerza ascendente de la flotación (F_B) y la fuerza de arrastre (F_D). Tanto la fuerza de gravedad como la de flotación se pueden calcular con la segunda ley de Newton, con la segunda igual al peso del fluido desplazado. Para flujo laminar, la fuerza de arrastre se puede calcular con la *ley de Stokes*,

$$F_D = 3\pi\mu d v$$

donde μ = viscosidad dinámica del fluido (N s/m^2), d = diámetro de la partícula (m) y v = velocidad de asentamiento de la partícula (m/s). Observe que la masa de la partícula se puede expresar como el producto del volumen de la partícula y su densidad ρ_s (kg/m^3), y la masa del fluido desplazado se puede calcular como el producto del volumen de la partícula y la densidad del fluido ρ (kg/m^3). El volumen de una esfera es $\pi d^3/6$. Además, el flujo laminar corresponde al caso en el que el número adimensional de Reynolds, Re, es menor de 1, donde Re = $\rho d v/\mu$.

a) Use un balance de fuerzas para la partícula a fin de desarrollar la ecuación diferencial para *dυ/dt* como función de *d*, ρ, ρ_s y μ.
b) En estado estable, use la ecuación para despejar la velocidad terminal de la partícula.
c) Utilice el resultado de *b*) para calcular la velocidad terminal de la partícula en m/s para una partícula esférica de limo que se sedimenta en agua: $d = 10$ μm, $\rho = 1$ g/cm³, $\rho_s = 2.65$ g/cm³ y $\mu = 0.014$ g /(cm · s).
d) Verifique si el flujo es laminar.
e) Use el método de Euler para calcular la velocidad desde $t = 0$ hasta 2^{-15} s, con $\Delta t = 2^{-18}$ s, dados los parámetros anteriores junto con la condición inicial: $\upsilon(0) = 0$.

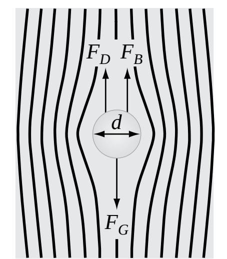


Figura P1.22

1.23 Como se describe en el problema 1.22, además de la fuerza descendente de la gravedad (peso) y del arrastre, el objeto que cae a través de un fluido también está sujeto a la fuerza de flotación, que es proporcional al volumen desplazado. Por ejemplo, para una esfera con diámetro *d*(m), el volumen de la esfera es $V = \pi d^3/6$, y su área proyectada es $A = \pi d^2/4$. La fuerza de flotación se puede calcular como $F_b = -\rho V_g$. En nuestra obtención de la ecuación (1.9), hemos despreciado la flotación porque es relativamente pequeña para un objeto como un paracaidista que se mueve en el aire. Sin embargo, para un fluido más denso como el agua, se vuelve más importante.
a) Obtenga una ecuación diferencial del mismo modo que la ecuación (1.9), pero incluya la fuerza de flotación y represente la fuerza de arrastre como se describe en el problema 1.21.
b) Reescriba la ecuación diferencial de *a*) para el caso especial de una esfera.
c) Use la ecuación desarrollada en *b*) para calcular la velocidad terminal (es decir, para el caso de estado estable). Use los siguientes valores de parámetros para la esfera que cae a través de agua: diámetro de la esfera = 1 cm; densidad de la esfera = 2 700 kg/m³; densidad del agua = 1 000 kg/m³ y $C_d = 0.47$.
d) Use el método de Euler con un tamaño de paso $\Delta t = 0.03125$ s para despejar numéricamente la velocidad desde $t = 0$ hasta 0.25 s, con velocidad inicial cero.

1.24 Como se ilustra en la figura P1.24, la deflexión hacia abajo *y* (m) de una viga en voladizo con una carga uniforme *w* (kg/m) se puede calcular como

$$y = \frac{w}{24EI}(x^4 - 4Lx^3 + 6L^2x^2)$$

donde x = distancia (m), E = módulo de elasticidad = 2×10^{11} Pa, I = momento de inercia = 3.25×10^{-4}m⁴, w = 10 000 N/m, y L = longitud = 4 m. Esta ecuación se puede diferenciar para obtener la pendiente de la deflexión hacia abajo como función de *x*:

$$\frac{dy}{dx} = \frac{w}{24EI}(4x^3 - 12Lx^2 + 12L^2x)$$

Si $y = 0$ en $x = 0$, use esta ecuación con el método de Euler ($\Delta x = 0.125$ m) para calcular la deflexión desde $x = 0$ hasta *L*. Desarrolle una gráfica de sus resultados junto con la solución analítica calculada con la primera ecuación.

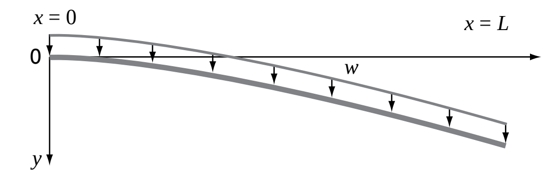


Figura P1.24
Viga en voladizo

1.25 Use el *principio de Arquímedes* y desarrolle un balance de fuerzas de estado estable para una bola esférica de hielo que flota en agua de mar (figura P1.25). El balance de fuerzas se debe expresar como polinomio de tercer grado (cúbico) en términos de la altura del casquete sobre el nivel del agua (*h*), la densidad del agua de mar (ρ_f), la densidad de la bola (ρ_s) y el radio de la bola (*r*).

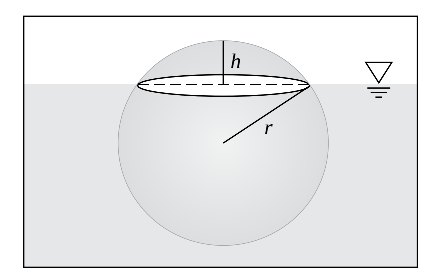


Figura P1.25

1.26 Además de su aplicación en fluidos, el *principio de Arquímedes* ha resultado útil en la geología, cuando se aplica a los sólidos en la corteza terrestre. La figura P1.26 ilustra un caso así, en donde una montaña cónica más ligera de granito "flota" sobre una capa más densa de basalto en la superficie de la tierra. Observe que la parte del cono que está debajo de la superficie se denomina frecuentemente *frústum*. Desarrolle un balance de fuerzas de estado de equilibrio para este caso en términos de los siguientes parámetros: densidad del basalto (ρ_b), densidad del granito (ρ_g), el radio de la base del cono (*r*) y la altura por encima de la superficie de la tierra, (h_1), y por debajo de dicha superficie (h_2).

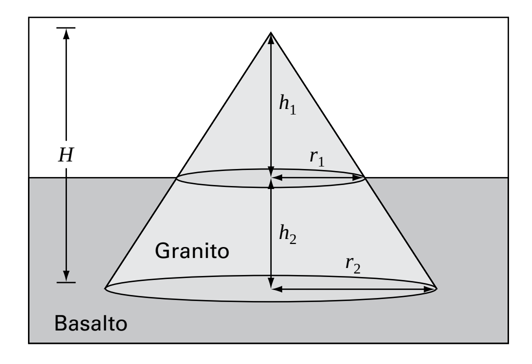


Figura P1.26

CAPÍTULO 2

Programación y software

En el capítulo 1, desarrollamos un modelo matemático a partir de la fuerza neta para predecir la velocidad de caída de un paracaidista. Este modelo tenía la forma de una ecuación diferencial,

$$\frac{dv}{dt} = g - \frac{c}{m}v$$

También vimos que se obtenía una solución de esta ecuación utilizando un método numérico simple, llamado método de Euler,

$$v_{i+1} = v_i + \frac{dv_i}{dt}\Delta t$$

Dada una condición inicial, se emplea esta ecuación repetidamente para calcular la velocidad como una función del tiempo. Sin embargo, para obtener una buena precisión sería necesario desarrollar muchos pasos pequeños. Hacerlo a mano sería muy laborioso y tomaría mucho tiempo; pero, con la ayuda de las computadoras, tales cálculos pueden realizarse fácilmente.

Por ende, nuestro siguiente objetivo consiste en descifrar cómo se hace esto. En el presente capítulo daremos una introducción al uso de la computadora como una herramienta para obtener soluciones de este tipo.

2.1 PAQUETES Y PROGRAMACIÓN

En la actualidad existen dos tipos de usuarios de software. Por un lado están aquellos que toman lo que se les da. Es decir, quienes se limitan a las capacidades que encuentran en el modo estándar de operación del software existente. Por ejemplo, resulta muy sencillo resolver un sistema de ecuaciones lineales o generar una gráfica con valores *x-y* con Excel o con MATLAB. Como este modo de operación por lo común requiere un mínimo esfuerzo, muchos de los usuarios adoptan este modo de operación. Además, como los diseñadores de estos paquetes se anticipan a la mayoría de las necesidades típicas de los usuarios, muchos de los problemas pueden resolverse de esta manera.

Pero, ¿qué pasa cuando se presentan problemas que están más allá de las capacidades estándar de dichas herramientas? Por desgracia, decir "lo siento, jefe, pero no lo sé hacer" no es algo aceptado en la mayoría de los círculos de la ingeniería. En tales casos usted tiene dos alternativas.

La primera sería buscar otro paquete y ver si sirve para resolver el problema. Ésta es una de las razones por las que quisimos usar tanto Excel como MATLAB en este libro. Como veremos, ninguno de los dos abarca todo y cada uno tiene sus ventajas. Sabiendo usar ambos, se amplía de forma notable el rango de problemas que pueden resolverse.

La segunda sería volverse un "potente usuario" si se aprende a escribir macros en Excel VBA[1] o archivos M (M-files) en MATLAB. ¿Y qué son éstos? No son más que programas computacionales que permiten ampliar la capacidad de estas herramientas. Como los ingenieros nunca se sentirán satisfechos al verse limitados por las herramientas, harán todo lo que sea necesario para resolver sus problemas. Una buena manera de lograrlo consiste en aprender a escribir programas en los ambientes de Excel y MATLAB. Además, las habilidades necesarias para crear macros o archivos M (M-files) son las mismas que se necesitan para desarrollar efectivamente programas en lenguajes como Fortran 90 o C.

El objetivo principal del capítulo es enseñarle cómo se hace esto. Sin embargo, supondremos que usted ya ha tenido contacto con los aspectos básicos de la programación y, por tal razón, desta-

[1] VBA son las siglas de Visual Basic for Applications.

caremos las facetas de la programación que afectan directamente su uso en la solución de problemas en ingeniería.

2.1.1 Programas computacionales

Los *programas computacionales* son únicamente conjuntos de instrucciones que dirigen a la computadora para realizar una cierta tarea. Dado que muchas personas escriben programas para una amplia gama de aplicaciones, la mayor parte de los lenguajes de alto nivel, como Fortran 90 o C, tienen amplias capacidades. Aunque habrá algunos ingenieros que usarán toda la amplia gama de capacidades, la mayoría sólo necesitará realizar los cálculos numéricos orientados a la ingeniería.

Visto desde esta perspectiva, podemos acotar la complejidad a unos cuantos tópicos de programación, que son:

- Representación de información sencilla (declaración de constantes, variables y tipos).
- Representación de información más compleja (estructuras de datos, arreglos y registros).
- Fórmulas matemáticas (asignación, reglas de prioridad y funciones intrínsecas).
- Entrada/Salida.
- Representación lógica (secuencia, selección y repetición).
- Programación modular (funciones y subrutinas).

Como suponemos que el lector ya ha tenido algún contacto con la programación, no dedicaremos mucho tiempo en las cuatro primeras áreas. A lo sumo, los ofrecemos como una lista de verificación que cubre lo que necesita saber para implementar los programas que siguen.

No obstante, sí dedicaremos algún tiempo a los dos últimos tópicos. Destacamos la representación lógica porque es el área que más influye en la coherencia y la comprensión de un algoritmo. Incluimos la programación modular porque también contribuye ampliamente a la organización del programa. Además, los módulos proveen de medios para archivar algoritmos útiles en un formato conveniente para aplicaciones subsecuentes.

2.2 PROGRAMACIÓN ESTRUCTURADA

En los comienzos de la computación, los programadores no daban mucha importancia a que sus programas fueran claros y fáciles de entender. Sin embargo, hoy se reconoce que escribir programas organizados y bien estructurados tiene muchas ventajas. Además de las ventajas obvias de tener un software más accesible para compartirlo, también ayuda a generar programas mucho más eficientes. Es decir, algoritmos bien estructurados, que son invariablemente mucho más fáciles de depurar y de probar, lo que resulta en programas que toman menos tiempo desarrollar, probar y actualizar.

Los científicos de la computación han estudiado sistemáticamente los factores y los procedimientos necesarios para desarrollar software de alta calidad de este tipo. En esencia la *programación estructurada* es un conjunto de reglas que prescriben buenos hábitos de estilo para el programador. Aunque la programación estructurada es bastante flexible para permitir considerable creatividad y expresión personal, sus reglas imponen suficientes restricciones para hacer que los programas resultantes sean muy superiores a sus versiones no estructuradas. En particular, el producto terminado es mucho más elegante y fácil de entender.

La idea clave detrás de la programación estructurada es que cualquier algoritmo numérico requiere tan sólo de tres estructuras de control fundamentales: secuencia, selección y repetición. Limitándonos a dichas estructuras el programa resultante será claro y fácil de seguir.

En los párrafos siguientes describiremos cada una de estas estructuras. Para mantener esta descripción de una manera general usaremos diagramas de flujo y pseudocódigo. Un *diagrama de flujo* es una representación visual o gráfica de un algoritmo. Un diagrama de flujo emplea una serie de bloques y flechas, cada una de las cuales representa un determinado paso u operación del algoritmo (figura 2.1). Las flechas representan el orden en el que se realizarán las operaciones.

No todas las personas relacionadas con la computación están de acuerdo en que los diagramas de flujo sean un esfuerzo productivo, de hecho, algunos programadores experimentados no usan los diagramas de flujo. Sin embargo, nosotros pensamos que existen tres buenas razones para estudiarlos. La primera es que sirven para expresar y comunicar algoritmos. La segunda es que aunque no

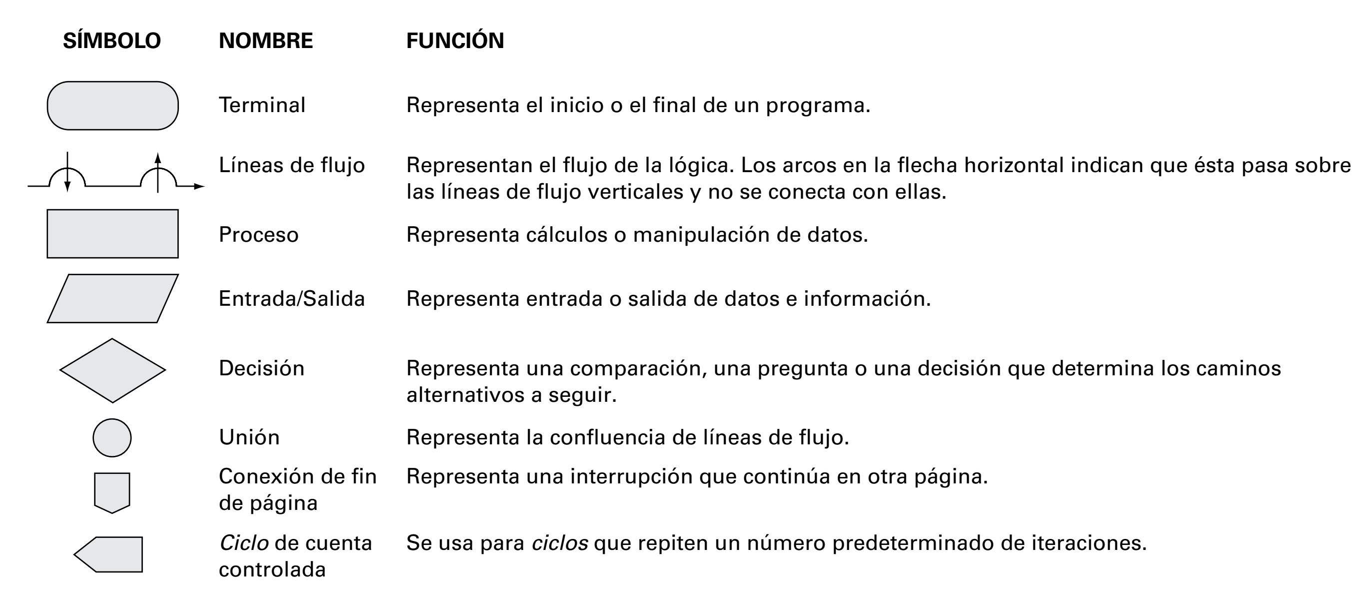


Figura 2.1
Símbolos usados en los diagramas de flujo.

se empleen de manera rutinaria, algunas veces resultarán útiles para planear, aclarar o comunicar la lógica del propio programa o del de otra persona. Por último, que es lo más importante para nuestros objetivos, son excelentes herramientas didácticas. Desde el punto de vista de la enseñanza, son los medios ideales para visualizar algunas de las estructuras de control fundamentales que se emplean en la programación.

Otra manera de expresar algoritmos, y que constituye un puente de unión entre los diagramas de flujo y el código de la computadora, es el *pseudocódigo.* En esta técnica se utilizan expresiones semejantes a las del código, en lugar de los símbolos gráficos del diagrama de flujo. En esta obra, para el pseudocódigo hemos adoptado algunas convenciones de estilo. Escribiremos con mayúsculas las palabras clave como IF, DO, INPUT, etc., mientras que las condiciones, pasos del proceso y tareas irán con una sangría. Además, los pasos del proceso se escribirán con una sangría. De esta manera las palabras clave forman un "sándwich" alrededor de los pasos para definir visualmente lo que abarca cada estructura de control.

Una ventaja del pseudocódigo es que con él resulta más fácil desarrollar un programa que con el diagrama de flujo. El pseudocódigo es también más fácil de modificar y de compartir con los demás. No obstante, los diagramas de flujo, debido a su forma gráfica, resultan a veces más adecuados para visualizar algoritmos complejos. Nosotros emplearemos diagramas de flujo con fines didácticos, y el pseudocódigo será el principal medio que usaremos para comunicar algoritmos relacionados con métodos numéricos.

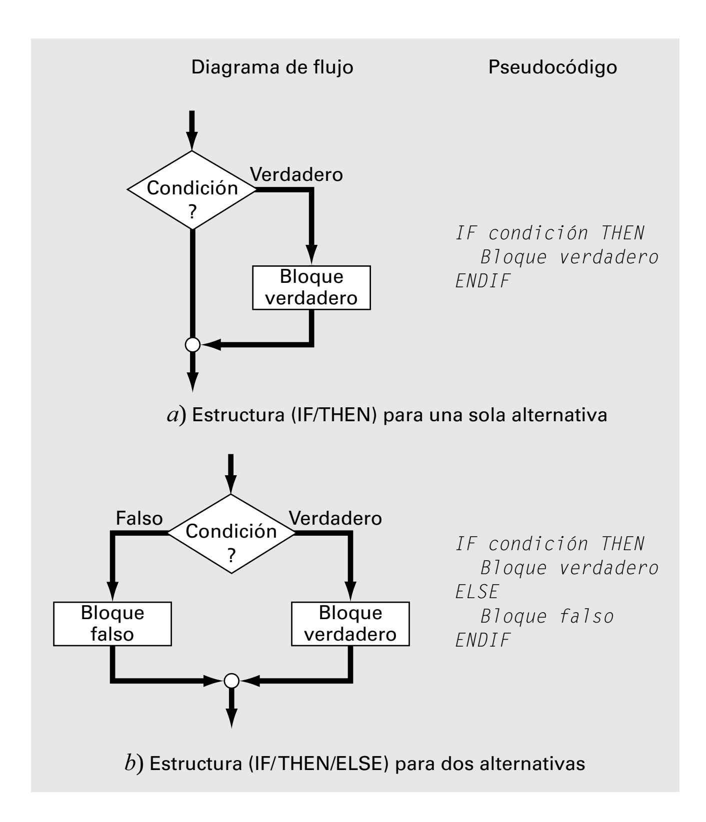


Figura 2.2
Diagrama de flujo y pseudocódigo para estructuras de selección simple. *a*) Selección con una alternativa (IF/THEN) y *b*) selección con dos alternativas (IF/THEN/ELSE).

2.2.1 Representación lógica

Secuencia La estructura secuencial expresa la trivial idea de que, a menos que se indique otra cosa, el código debe rea-

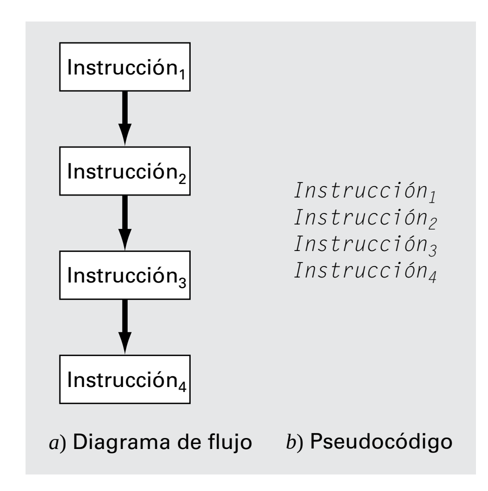


Figura 2.3
a) Diagrama de flujo y *b*) pseudocódigo para una estructura secuencial.

lizarse instrucción por instrucción. Como en la figura 2.2, la estructura se puede expresar de manera general como un diagrama de flujo o como un pseudocódigo.

Selección En contraste con el paso por paso de la estructura secuencial, la selección nos ofrece un medio de dividir el flujo del programa en ramas considerando el resultado de una condición lógica. La figura 2.3 muestra las dos principales formas de hacer esto.

La decisión ante una sola alternativa, o estructura *IF*/*THEN* (figura 2.3*a*), nos permite una desviación en el flujo del programa si una condición lógica es verdadera. Si esta condición es falsa no ocurre nada y el programa continúa con la indicación que se encuentra después del ENDIF. La decisión ante dos alternativas, o estructura *IF*/*THEN*/*ELSE* (figura 2.3*b*), se comporta de la misma manera si la condición es verdadera; sin embargo, si la condición es falsa, el programa realiza las instrucciones entre el ELSE y el ENDIF.

Aunque las estructuras IF/THEN e IF/THEN/ELSE son suficientes para construir cualquier algoritmo numérico, por lo común también se usan otras dos variantes. Suponga que el ELSE de un IF/THEN/ELSE contiene otro IF/THEN. En tales casos el ELSE y el IF se pueden combinar en la estructura *IF/THEN/ELSEIF* que se muestra en la figura 2.4*a*.

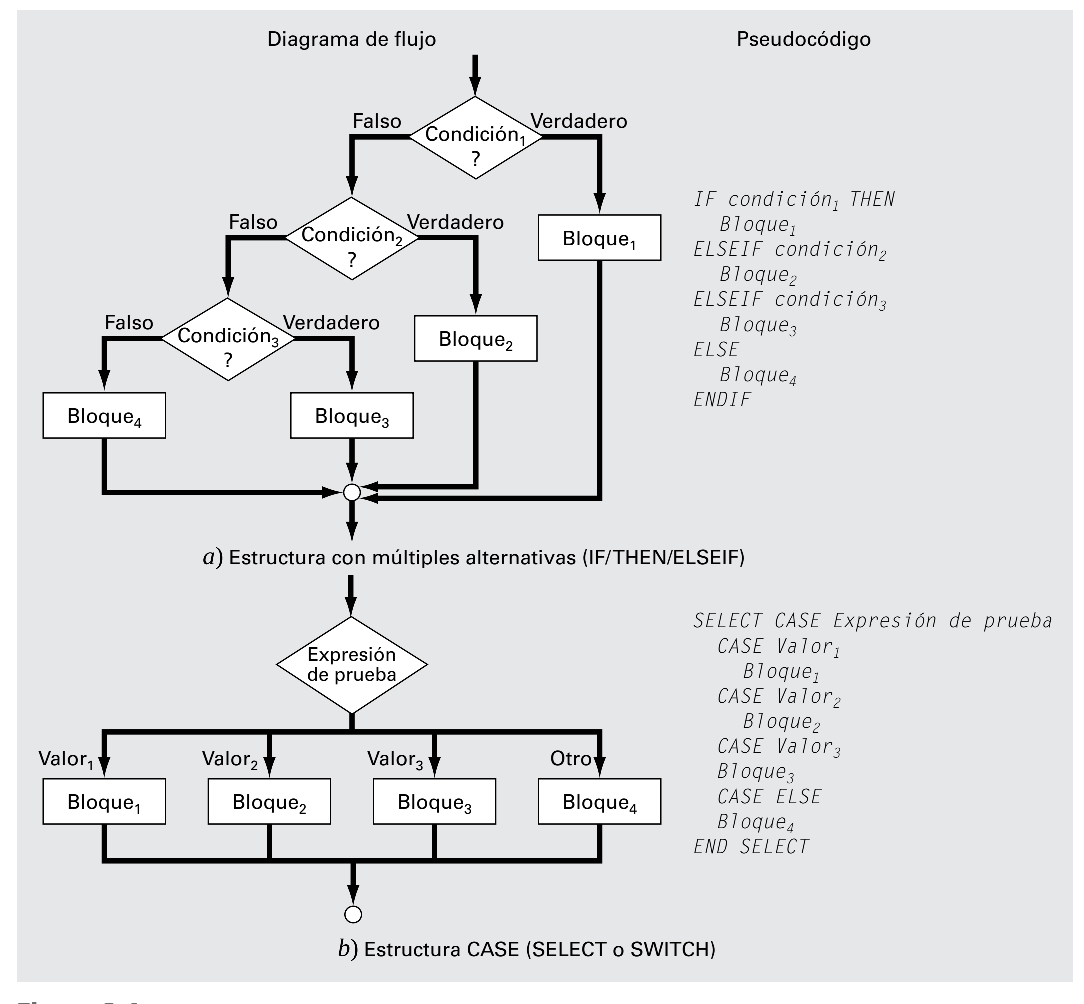


Figura 2.4
Diagrama de flujo y pseudocódigo para construcciones de selección o ramificación. *a*) Selección de múltiples alternativas (IF/THEN/ELSEIF) y *b*) Construcción CASE.

Observe que en la figura 2.4*a* hay una cadena o "cascada" de decisiones. La primera es una instrucción IF y cada una de las decisiones sucesivas es un ELSEIF. Siguiendo la cadena hacia abajo, la primera condición que resulte verdadera ocasionará una desviación a su correspondiente bloque de código, seguida por la salida de la estructura. Al final de la cadena de condiciones, si todas las condiciones resultaron falsas, se puede adicionar un bloque ELSE opcional.

La estructura *CASE* es una variante de este tipo de toma de decisiones (figura 2.4*b*). En lugar de probar condiciones individuales, las ramificaciones dependen del valor de una sola *expresión de prueba*. Según sea su valor, se presentarán diferentes bloques de código. Además, si la expresión no toma ninguno de los valores previstos, se puede proponer un bloque opcional (CASE ELSE).

Repetición La repetición nos proporciona una manera de llevar a cabo instrucciones repetidamente. Las estructuras resultantes, llamadas *loops* o ciclos, se presentan en dos formas distintas que se diferencian por la manera en que terminan.

El primer tipo, y el fundamental, es el llamado *ciclo de decisión* debido a que termina basándose en el resultado de una condición lógica. La figura 2.5 muestra el tipo general de ciclo de decisión, la *construcción DOEXIT,* también llamada *ciclo de interrupción* (*break ciclo*). Esta estructura realiza repeticiones hasta que una condición lógica resulte verdadera.

En esta estructura no es necesario tener dos bloques. Cuando se omite el primer bloque, a la estructura se le suele llamar *ciclo de preprueba* porque la prueba lógica se realiza antes de que ocurra algo. Si se omite el segundo bloque, se le llama *ciclo posprueba*. Al caso general (figura 2.5), en el que se incluyen los dos bloques, se le llama *ciclo de prueba intermedia* (*midtest ciclo*).

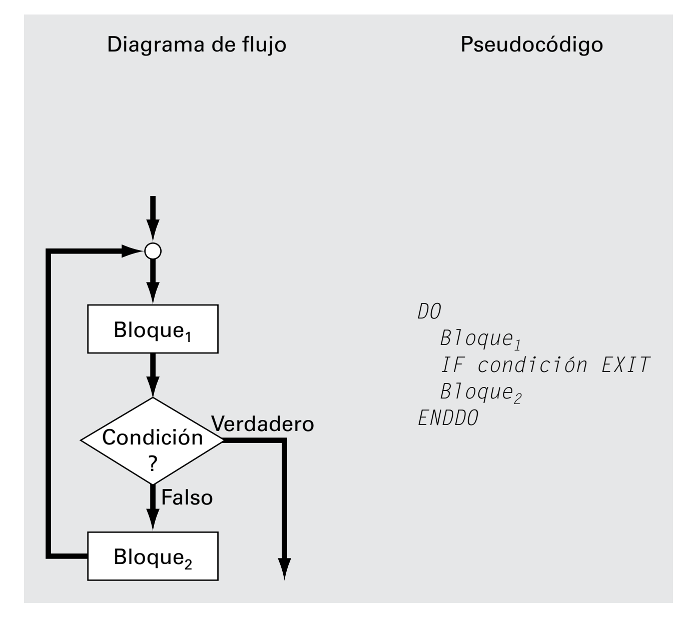


Figura 2.5
Ciclo DOEXIT o de interrupción.

Hay que hacer notar que el ciclo DOEXIT fue introducido en Fortran 90 para tratar de simplificar los ciclos de decisión. Esta estructura de control es parte estándar del lenguaje VBA de macros en Excel; pero no forma parte estándar de C o de MATLAB, que usan la estructura llamada WHILE. Como nosotros consideramos superior a la estructura DOEXIT, la hemos adoptado en este libro como la estructura de ciclo de decisión. Para que nuestros algoritmos se realicen tanto en MATLAB como en Excel, mostraremos más adelante, en este capítulo (véase la sección 2.5), cómo simular el ciclo de interrupción usando la estructura WHILE.

Al ciclo de interrupción que se presenta en la figura 2.5 se le llama ciclo lógico porque termina a causa de una condición lógica. Por otro lado, se tiene el *ciclo controlado por contador* o *ciclo DOFOR* (figura 2.6) que realiza un número determinado de repeticiones o iteraciones.

El ciclo controlado por contador funciona como sigue. El índice (representado por *i* en la figura 2.6) es una variable a la que se le da un valor *inicial*. El programa prueba si el *índice* es menor o igual al valor final, *fin*. Si es así, entonces ejecuta el cuerpo del ciclo y vuelve al DO. Cada vez que encuentra el ENDDO el *índice* se incrementa automáticamente con el valor definido por el incremento. De manera que el índice actúa como un contador. Cuando el *índice* es mayor que el valor final (*fin*), la computadora sale automáticamente del ciclo y transfiere el control a la línea que sigue después del ENDDO. Observe que casi en todos los lenguajes de programación, incluyendo Excel y MATLAB, si se omite el *incremento*, la computadora supone que éste es igual a 1.[2]

Los algoritmos numéricos que se describen en las páginas siguientes se desarrollarán usando únicamente las estructuras presentadas en las figuras 2.2 a 2.6. El ejemplo siguiente presenta el método básico para desarrollar un algoritmo que determine las raíces de la ecuación cuadrática.

[2] Se puede usar incremento (decremento) negativo, en cuyo caso el ciclo termina cuando el índice es menor que el valor final.

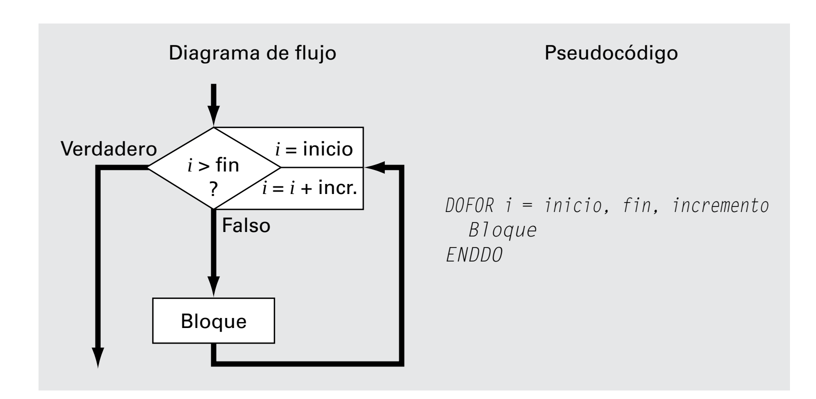


Figura 2.6
Construcción controlada por conteo o construcción DOFOR.

EJEMPLO 2.1 Algoritmo para las raíces de la ecuación cuadrática

Planteamiento del problema Las raíces de una ecuación cuadrática

$$ax^2 + bx + c = 0$$

se determinan mediante la fórmula cuadrática,

$$\begin{matrix} x_1 \\ x_2 \end{matrix} = \frac{-b \pm \sqrt{|b^2 - 4ac|}}{2a} \tag{E2.1.1}$$

Desarrolle un algoritmo que haga lo siguiente:

Paso 1: Pida al usuario los coeficientes *a*, *b* y *c*.
Paso 2: Realice las operaciones de la fórmula cuadrática previendo todas las eventualidades (como, por ejemplo, evitar la división entre cero y permitir raíces complejas).
Paso 3: Dé la solución, es decir, los valores de *x*.
Paso 4: Dé al usuario la opción de volver al paso 1 y repetir el proceso.

Solución Para desarrollar el algoritmo usaremos un método que va de lo general a lo particular (método *top-down*). Esto es, iremos refinando cada vez más el algoritmo en lugar de detallar todo a la primera vez.

Para hacerlo, supongamos, para el presente ejemplo, la fórmula cuadrática es infalible sin importar los coeficientes (obviamente no es cierto, pero por ahora basta). Un algoritmo estructurado para realizar la tarea es

```
DO
    INPUT a, b, c
    r1 = (–b + SQRT (b² – 4ac))/(2a)
    r2 = (–b – SQRT (b² – 4ac))/(2a)
    DISPLAY r1, r2
    DISPLAY '¿Repetir? Conteste sí o no'
    INPUT respuesta
    IF respuesta = 'no' EXIT
ENDDO
```

La construcción DOEXIT se utiliza para repetir el cálculo de la ecuación cuadrática siempre que la condición sea falsa. La condición depende del valor de la variable de tipo carácter *respuesta*. Si *respuesta* es igual a 'sí' entonces se llevan a cabo los cálculos. Si no es así, si *respuesta* es igual a 'no', el ciclo termina. De esta manera, el usuario controla la terminación mediante el valor de *respuesta*.

Ahora bien, aunque el algoritmo anterior funcionará bien en ciertos casos, no es infalible. El algoritmo quizá no funcione para algunos valores de las variables. Esto es:

- Si $a = 0$ se presentará inmediatamente un problema debido a la división entre cero. Si inspeccionamos cuidadosamente la ecuación (2.1) veremos que aquí se pueden presentar dos casos:
  Si $b \neq 0$, la ecuación se reduce a una ecuación lineal con una raíz real, $-c/b$.
  Si $b = 0$, entonces no hay solución. Es decir, el problema es trivial.
- Si $a \neq 0$, entonces, según sea el valor del discriminante, $d = b^2 - 4ac$, se pueden presentar también dos casos:
  Si $d \geq 0$, habrá dos raíces reales.
  Si $d < 0$, habrá dos raíces complejas.

Observe cómo hemos dejado una sangría adicional para hacer resaltar la estructura de decisión que subyace a las matemáticas. Esta estructura se traduce, después, en un conjunto de estructuras IF/THEN/ELSE acopladas que se pueden insertar en la parte con los comandos sombreados en el código anterior, obteniéndose finalmente el algoritmo:

```
DO
  INPUT a, b, c
  r1 = 0: r2 = 0: i1 = 0: i2 = 0
  IF a = 0 THEN
     IF b ≠ 0 THEN
        r1 = -c/b
     ELSE
        DISPLAY "Solución trivial"
     ENDIF
  ELSE
     discr = b² - 4 * a * c
     IF discr ≥ 0 THEN
        r1 = (-b + Sqrt(discr))/(2 * a)
        r2 = (-b - Sqrt(discr))/(2 * a)
     ELSE
        r1 = -b/(2 * a)
        r2 = r1
        i1 = Sqrt(Abs(discr))/(2 * a)
        i2 = -i1
     ENDIF
  ENDIF
  DISPLAY r1, r2, i1, i2
  DISPLAY '¿Repetir? Conteste sí o no'
  INPUT respuesta
  IF respuesta = 'no' EXIT
ENDDO
```

El método que se utilizó en el problema anterior puede emplearse para desarrollar un algoritmo para el problema del paracaidista. Recordemos que, dadas la condición inicial para tiempo y velocidad, el problema consistía en resolver de manera iterativa la fórmula

$$v_{i+1} = v_i + \frac{dv_i}{dt}\Delta t \tag{2.1}$$

Como sabemos, para lograr una buena precisión será necesario emplear incrementos pequeños. Por lo que será necesario emplear la fórmula repetidas veces, desde el tiempo inicial hasta el tiempo final. En consecuencia, un algoritmo para resolver este problema estará basado en el uso de un ciclo.

Supongamos, por ejemplo, que empezamos los cálculos en $t = 0$ y queremos predecir la velocidad en $t = 4$ s con incrementos de tiempo $\Delta t = 0.5$ s. Entonces tendremos que aplicar la ecuación (2.1) ocho veces, esto es,

$$n = \frac{4}{0.5} = 8$$

donde *n* es el número de iteraciones del ciclo. Como este número es exacto, es decir, esta división nos da un número entero, podemos usar como base del algoritmo un ciclo controlado por contador. A continuación damos un ejemplo de pseudocódigo.

```
g = 9.81
INPUT cd, m
INPUT ti, vi, tf, dt
t = ti
v = vi
n = (tf – ti) / dt
DOFOR i = 1 TO n
   dvdt = g – (cd / m) * v
   v = v + dvdt * dt
   t = t + dt
ENDDO
DISPLAY v
```

Aunque este esquema es fácil de programar, no está completo. Sólo funcionará si el intervalo es divisible exactamente entre el incremento.[3] Para tomar en cuenta el otro caso, en el pseudocódigo anterior, en lugar del área sombreada se puede usar un ciclo de decisión. El resultado es:

```
g = 9.81
INPUT cd, m
INPUT ti, vi, tf, dt
t = ti
v = vi
h = dt
DO
   IF t + dt > tf THEN
      h = tf – t
  ENDIF
   dvdt = g – (cd / m) * v
   v = v + dvdt * h
   t = t + h
   IF t ≥ tf EXIT
ENDDO
DISPLAY v
```

Al introducir el ciclo, usamos la estructura IF/THEN para probar si el valor $t + dt$ nos lleva más allá del final del intervalo. Si no es así, lo cual comúnmente será el caso al principio, no hacemos nada. De lo contrario, necesitaremos reducir el intervalo haciendo el tamaño de incremento *h* igual a $tf – t$. Así, garantizamos que el paso siguiente caiga precisamente en *tf*. Después de hacer este paso final, el ciclo terminará, debido a que $t \geq tf$ será verdadero.

Observe que antes de entrar en el ciclo hemos asignado el valor del incremento, *dt*, a otra variable, *h*. Creamos esta variable con el objeto de que nuestra rutina no cambie el valor de *dt* cuando tengamos que reducir el incremento. Hacemos esto anticipándonos a que tengamos que usar el valor original de *dt* en algún otro lado, en el caso de que este programa sea parte de otro programa mayor.

Hay que destacar que este algoritmo aún no está terminado. Puede ser, por ejemplo, que el usuario dé por error un incremento que sea mayor que el intervalo, como por ejemplo, $tf – ti = 5$ y $dt = 20$. Entonces, habrá que poner, en el programa, instrucciones que detectan tales errores y que el usuario pueda corregirlos.

---

[3] Este problema se combina con el hecho de que las computadoras usan internamente, para la representación de números, la base 2. En consecuencia, algunos números que aparentemente son divisibles no dan exactamente un entero cuando la división se hace en una computadora. De esto hablaremos en el capítulo 3.

2.3 PROGRAMACIÓN MODULAR

Imaginemos qué difícil sería estudiar un libro que no tuviera capítulos, ni secciones, ni párrafos. Dividir una tarea o una materia complicada en partes más accesibles es una manera de hacerla más fácil. Siguiendo esta misma idea, los programas de computación se dividen en subprogramas más pequeños, o módulos que pueden desarrollarse y probarse por separado. A esta forma de trabajar se le llama *programación modular.*

La principal cualidad de los módulos es que son tan independientes y autosuficientes como sea posible. Además, en general, están diseñados para llevar a cabo una función específica y bien definida, y tienen un punto de entrada y un punto de salida. Los módulos a menudo son cortos (por lo general 50 a 100 instrucciones) y están bien enfocados.

En los lenguajes estándar de alto nivel como Fortran 90 y C, el principal elemento de programación usado para representar módulos es el procedimiento. Un procedimiento es un conjunto de instrucciones para computadora que juntas realizan una tarea dada. Se emplean comúnmente dos tipos de procedimientos: *funciones* y *subrutinas*. Las primeras normalmente dan un solo resultado, mientras que las últimas dan varios.

```
FUNCTION Euler(dt, ti, tf, yi)
t = ti
y = yi
h = dt
DO
   IF t + dt > tf THEN
      h = tf − t
   ENDIF
   dydt = dy(t, y)
   y = y + dydt * h
   t = t + h
   IF t ≥ tf EXIT
ENDDO
Euler = y
END EULER
```

Figura 2.7
Pseudocódigo para una función que resuelve una ecuación diferencial usando el método de Euler.

Además, hay que mencionar que gran parte de la programación relacionada con paquetes de software como Excel y MATLAB implica el desarrollo de subprogramas. Así, los macros de Excel y las funciones de MATLAB están diseñadas para recibir información, llevar a cabo un cálculo y dar un resultado. De manera que el pensamiento modular también es consistente con la manera en que se programa en ambientes de paquetes.

La programación modular tiene diversas ventajas. El uso de unidades pequeñas e independientes hace que la lógica subyacente sea más fácil de seguir y de entender, tanto para el que desarrolla el módulo como para el usuario. Se facilita el desarrollo debido a que se puede perfeccionar cada módulo por separado. En proyectos grandes, varios programadores pueden trabajar por separado las diferentes partes individuales. En el diseño modular también la depuración y la prueba de un programa se simplifican debido a que los errores se pueden encontrar con facilidad. Por último, es más sencillo el mantenimiento y la modificación del programa. Esto se debe principalmente a que se pueden desarrollar nuevos módulos que realicen tareas adicionales e incorporarlos en el esquema coherente y organizado que ya se tiene.

Aunque todas esas ventajas son razones suficientes para usar módulos, la razón más importante, relacionada con la solución de problemas numéricos en ingeniería, es que permiten tener una biblioteca de módulos útiles para posteriores usos en otros programas. Ésta será la filosofía de la presente obra: todos los algoritmos serán presentados como módulos.

El procedimiento anterior se ilustra en la figura 2.7 que muestra una función desarrollada para usar el método de Euler. Observe que esa función y las versiones previas difieren en cómo manipulan la entrada y la salida (input/output). En las versiones anteriores directamente la entrada viene (mediante el INPUT) del usuario, y la salida va (mediante el DISPLAY) al usuario. En la función, se le da la entrada a ésta mediante su lista de argumentos FUNCTION

```
Function Euler(dt, ti, tf, yi)
```

y la salida es regresada mediante una asignación

```
y = Euler(dt, ti, tf, yi)
```

Observe, además, lo general que se ha vuelto esta rutina. No se hace para nada referencia al caso específico del paracaidista. Por ejemplo, dentro de la función, en lugar de llamar a la variable dependiente v, de velocidad, se le nombra y, de manera más general. Asimismo, note que la derivada no se calcula mediante una ecuación explícita dentro de la función. En lugar de ello se llama a otra función *dy* para calcularla, lo cual indica el hecho de que podemos usar esta función en muchos problemas distintos, además de encontrar la velocidad del paracaidista.

2.4 EXCEL

Excel es una hoja de cálculo producida por Microsoft, Inc. Las hojas de cálculo son un tipo especial de software para matemáticas que permite al usuario ingresar y realizar cálculos en renglones y columnas de datos. Como tales, son una versión computarizada de una gran hoja de contabilidad en la que se lleva a cabo una gran cantidad de cálculos interrelacionados. Puesto que cuando se modifica un valor de la hoja, hay que actualizar todos los cálculos, las hojas de cálculo son ideales para hacer análisis del tipo "¿y qué pasa si...?"

Excel cuenta con varios recursos numéricos interconstruidos como resolución de ecuaciones, ajuste de curvas y optimización. Incluye también VBA como un lenguaje de macro que sirve para hacer cálculos numéricos. Por último, tiene varias herramientas para la visualización como gráficas y superficies de tres dimensiones, que son un valioso complemento para el análisis numérico. En esta sección mostraremos cómo se utilizan estos recursos en la solución del problema del paracaidista.

Para ello, construimos primero una hoja de cálculo sencilla. Como se ve abajo, el primer paso consiste en colocar números y letras o palabras en las celdas de la hoja de cálculo.

	A	B	C	D
1	**Problema del paracaidista**			
2				
3	**m**	**68.1**	**kg**	
4	**cd**	**12.5**	**kg/s**	
5	**dt**	**0.1**	**s**	
6				
7	**t**	**vnum (m/s)**	**vanal (m/s)**	
8	**0**	**0.000**		
9	**2**			

Antes de escribir un programa de macro para calcular el valor numérico, podemos facilitar el trabajo subsecuente dando nombres a los valores de los parámetros. Para esto, seleccione las celdas A3:B5 (la manera más fácil de hacerlo es mover el ratón hasta A3, mantener oprimido el botón izquierdo del ratón y arrastrarlo hasta B5). Después seleccione, del menú,

```
Insert Name Create Left column OK
```

Para verificar que todo haya funcionado correctamente, seleccione la celda B3 y verifique que aparezca la etiqueta "m" en la casilla del nombre (casilla que se encuentra en el lado izquierdo de la hoja, justo debajo de las barras del menú).

Muévase hasta la celda C8 e introduzca la solución analítica (ecuación 1.9),

```
=9.8*m/cd*(1-exp(-cd/m*A8))
```

Al introducir esta fórmula debe aparecer el valor 0 en la celda C8. Después copie la fórmula a la celda C9 para obtener 16.405 m/s.

Todo lo anterior es típico del uso estándar de Excel. Hecho esto, podría, por ejemplo, cambiar los valores de los parámetros y observar cómo se modifica la solución analítica.

Ahora mostraremos cómo se usan las macros de VBA para extender los recursos estándar. En la figura 2.8 se da una lista que contiene, para cada una de las estructuras de control dadas en la sección anterior (figuras 2.2 a 2.6), el pseudocódigo junto con el código VBA de Excel. Observe que, aunque los detalles difieren, la estructura del pseudocódigo y la del código VBA son idénticas.

Ahora podemos usar algunas de las construcciones dadas en la figura 2.8 para escribir una función de macro que calcule la velocidad. Para abrir VBA seleccione[4]

```
Tools Macro Visual Basic Editor
```

[4] ¡La combinación de las teclas Alt-F11 es más rápida!

Una vez dentro del *Visual Basic Editor* (VBE), seleccione

```
Insert Module
```

y se abrirá una nueva ventana para código. La siguiente función en VBA se puede obtener directamente del pseudocódigo de la figura 2.7. Escriba la función dentro de la ventana del código.

```
Option Explicit
Function Euler(dt, ti, tf, yi, m, cd)

Dim h As Double, t As Double, y As Double, dydt As Double
t = ti
y = yi
h = dt
Do
   If t + dt > tf Then
     h = tf – t
   End If
   dydt = dy(t, y, m, cd)
   y = y + dydt * h
   t = t + h
   If t >= tf Then Exit Do
Loop
Euler = y
End Function
```

Compare esta macro con el pseudocódigo de la figura 2.7 y vea que son muy similares. Observe también cómo la lista de argumentos de la función se hizo más larga al incluir los parámetros necesarios para el modelo de la velocidad del paracaidista. La velocidad obtenida, v, pasa a la hoja de cálculo mediante el nombre de la función.

Note también cómo, para calcular la derivada, hemos usado otra función. Ésta se puede introducir en el mismo módulo tecleándola directamente debajo de la función Euler,

```
Function dy(t, v, m, cd)
Const g As Double = 9.81
dy = g – (cd / m) * v
End Function
```

El paso final consiste en volver a la hoja de cálculo y llamar a la función introduciendo la siguiente expresión en la celda B9

```
=Euler(dt,A8,A9,B8,m,cd)
```

El resultado de la integración numérica, 16.531, aparecerá en la celda B9.

Vamos a ver qué ha pasado aquí. Cuando usted da la función en la celda de la hoja de cálculo, los parámetros pasan al programa VBA, donde se realizan los cálculos y, después, el resultado regresa a la celda. En efecto, el lenguaje de macros VBA le permite usar Excel como mecanismo de entradas y salidas (input/output). Esta característica resulta de mucha utilidad.

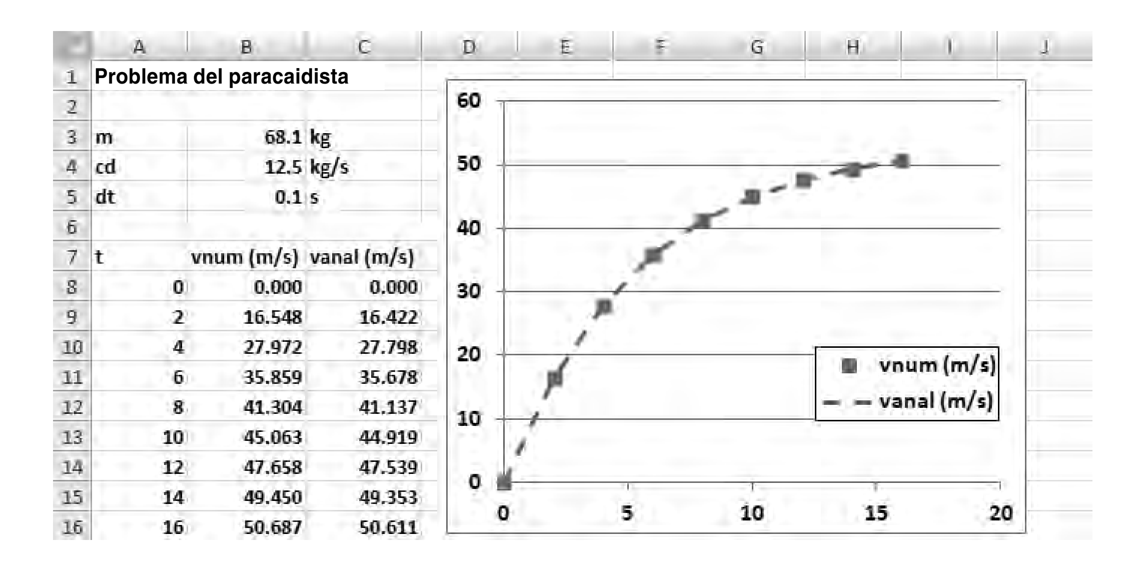
Por ejemplo, ahora que ya tiene todos los cálculos, puede jugar con ellos. Suponga que el paracaidista fuera mucho más pesado, digamos, m = 100 kg (alrededor de 220 libras). Introduzca 100 en la celda B3 y la hoja de cálculo se modificará de inmediato mostrando el valor 17.438 en la celda B9. Cambie la masa nuevamente a 68.1 kg y el resultado anterior, 16.531 reaparecerá de forma automática en la celda B9.

Ahora vayamos un poco más adelante dando algunos valores más para el tiempo. Introduzca los números 4, 6, …, 16 en las celdas A10 a A16. Después copie las fórmulas de las celdas B9:C9 hacia abajo en los renglones 10 a 16. Observe cómo el programa VBA calcula correctamente los resultados numéricos en cada uno de los nuevos renglones. (Para verificar esto cambie el valor de *dt* por 2 y compare los resultados con los cálculos a mano obtenidos anteriormente, en el ejemplo 1.2.) Para mejorar la presentación se pueden graficar los resultados en un plano *x-y* usando Excel Chart Wizard.

a) **Pseudocódigo**	*b*) **Excel VBA**
IF/THEN:	
IF condición THEN	If b <> 0 Then
Bloque verdadero	r1 = -c / b
ENDIF	End If
IF/THEN/ELSE:	
IF condición THEN	If a < 0 Then
Bloque verdadero	b = Sqr(Abs(a))
ELSE	Else
Bloque falso	b = Sqr(a)
ENDIF	End If
IF/THEN/ELSEIF:	
IF $condición_1$ *THEN*	If class = 1 Then
$Bloque_1$	x = x + 8
ELSEIF $condición_2$	ElseIf class < 1 Then
$Bloque_2$	x = x - 8
ELSEIF $condición_3$	ElseIf class < 10 Then
$Bloque_3$	x = x - 32
ELSE	Else
$Bloque_4$	x = x - 64
ENDIF	End If
CASE:	
SELECT CASE Expresión de prueba	Select Case a + b
CASE $Valor_1$	Case Is < -50
$Bloque_1$	x = -5
CASE $Valor_2$	Case Is < 0
$Bloque_2$	x = -5 - (a + b) / 10
CASE $Valor_3$	Case Is < 50
$Bloque_3$	x = (a + b) / 10
CASE ELSE	Case Else
$Bloque_4$	x = 5
END SELECT	End Select
DOEXIT:	
DO	Do
$Bloque_1$	i = i + 1
IF condición EXIT	If i >= 10 Then Exit Do
$Bloque_2$	j = i*x
ENDIF	Loop
LOOP CONTROLADO POR CONTADOR:	
DOFOR i = inicio, fin, incremento	For i = 1 To 10 Step 2
Bloque	x = x + i
ENDDO	Next i

Figura 2.8
Estructuras de control fundamentales en *a*) pseudocódigo y *b*) VBA de Excel.

Abajo se muestra la hoja de cálculo resultante. Hemos creado una valiosa herramienta para la solución de problemas. Puede realizar un análisis de sensibilidad cambiando los valores de cada uno de los parámetros. Cada vez que se introduce un nuevo valor, se actualizarán automáticamente los cálculos y la gráfica. Tal característica de interactividad es lo que hace tan potente a Excel. No obstante, se debe reconocer que resolver este problema dependerá de la habilidad para escribir el macro en VBA.



	A	B	C
1	Problema del paracaidista		
2			
3	m	68.1	kg
4	cd	12.5	kg/s
5	dt	0.1	s
6			
7	t	vnum (m/s)	vanal (m/s)
8	0	0.000	0.000
9	2	16.548	16.422
10	4	27.972	27.798
11	6	35.859	35.678
12	8	41.304	41.137
13	10	45.063	44.919
14	12	47.658	47.539
15	14	49.450	49.353
16	16	50.687	50.611

La combinación del ambiente de Excel con el lenguaje de programación VBA nos abre un mundo de posibilidades para la solución de problemas en ingeniería. En los capítulos siguientes ilustraremos cómo se logra esto.

2.5 MATLAB

MATLAB es el principal producto de software de The MathWorks, Inc., fundada por los analistas numéricos Cleve Moler y John N. Little. Como su nombre lo indica, MATLAB se desarrolló originalmente como un *lab*oratorio para *mat*rices. Hoy, el elemento principal de MATLAB sigue siendo la matriz. La manipulación matemática de matrices se ha realizado muy adecuadamente en un ambiente interactivo fácil de utilizar. A esta manipulación matricial, MATLAB agrega varias funciones numéricas, cálculos simbólicos y herramientas para visualización. En consecuencia, la versión actual representa un ambiente computacional bastante amplio.

MATLAB tiene diferentes funciones y operadores que permiten la adecuada realización de los métodos numéricos que aquí desarrollamos. Éstos se describirán con detalle en los capítulos siguientes. Además, se pueden escribir programas como los llamados archivos M (*M-files*) que sirven para realizar cálculos numéricos. Vamos a explorar cómo funciona.

Primero, usted se dará cuenta de que el uso normal de MATLAB está estrechamente relacionado con la programación. Supongamos, por ejemplo, que queremos determinar la solución analítica al problema del paracaidista, lo cual haríamos con los siguientes comandos de MATLAB

```
>> g=9.81;
>> m=68.1;
>> cd=12.5;
>> tf=2;
>> v=g*m/cd*(1-exp(-cd/m*tf))
```

obteniéndose como resultado

```
v =
   16.4217
```

La secuencia de comandos es como la secuencia de instrucciones en un lenguaje de programación típico.

Pero, ¿qué ocurre si usted se quiere desviar de la estructura secuencial? Aunque hay algunos caminos bien definidos para establecer recursos no secuenciales en el modo estándar de comandos, para introducir decisiones y ciclos, lo mejor es crear un documento de MATLAB al que se le llama archivo M (*M-file*). Para hacer esto haga clic en

```
File New Script
```

y se abrirá una ventana nueva con el encabezado “MATLAB Editor/Debugger”. En esta ventana usted puede escribir y editar programas en MATLAB. Escriba ahí el código siguiente:

```
g=9.81;
m=68.1;
cd=12.5;
tf=2;
v=g*m/cd*(1-exp(-cd/m*tf))
```

Obsérvese que los comandos se escriben exactamente en la misma forma en que se haría en el extremo frontal de MATLAB. Guarde el programa con el mismo nombre: analpara. MATLAB agregará en forma automática la extensión .m para denotar que se trata de un archivo M: analpara.m.

Para correr el programa, se debe regresar al modo de comando. La forma más directa de efectuar esto consiste en hacer clic en el botón “MATLAB Command Window” que se encuentra en la barra de tareas (que por lo general está en la parte inferior de la pantalla).

Ahora, el programa se puede correr al hacer clic en el archivo M, analpara, que debe parecerse a lo siguiente:

```
>> analpara
```

Si usted ha hecho todo en forma correcta, MATLAB debe responder con la respuesta correcta:

```
v =
   16.4217
```

Ahora, un problema con lo anterior es que está preparado para calcular sólo un caso. El lector lo puede hacer más flexible si hace que el usuario introduzca algunas de las variables. Por ejemplo, suponga que desea evaluar el efecto de la masa sobre la velocidad a los 2 s. Para hacer esto, el archivo M podría reescribirse como sigue:

```
g=9.81;
m=input('masa (kg):');
cd=12.5;
tf=2;
v=g*m/cd*(1-exp(-cd/m*tf))
```

Guarde esto con el nombre de analpara2.m. Si escribió analpara2 mientras se encontraba en el modo de comando, la línea mostrará lo que sigue:

```
masa (kg):
```

Entonces, el usuario introduce un valor como 100, y el resultado aparecerá como:

```
v =
   17.3597
```

Ahora, debe quedar bastante claro cómo se puede programar una solución numérica por medio de un archivo M. A fin de hacerlo, primero debemos entender la manera en que MATLAB maneja las estructuras lógica y de lazo (ciclos o *loops*). En la figura 2.9 se enlista el pseudocódigo junto con el código de MATLAB para todas las estructuras de control, con base en la sección anterior. Aunque las estructuras del pseudocódigo y el código MATLAB son muy similares, existen algunas diferencias pequeñas que deben destacarse.

En especial, observe cómo hemos expresado la estructura DOEXIT. En lugar del DO usamos el WHILE(1). Como MATLAB interpreta al número 1 como correspondiente a “verdadero”, esta instrucción se repetirá indefinidamente de la misma manera que el DO. El *loop* termina con un comando de *interrupción* (*break*), el cual transfiere el control a la instrucción que se encuentra a continuación, de la instrucción *end* que termina el ciclo.

También hay que observar que los parámetros del lazo controlado por contador están ordenados de modo diferente. Para el pseudocódigo, los parámetros del lazo están especificados como *`start, finish, step`*. Para MATLAB, los parámetros están ordenados como `start:step:finish`.

Ahora el siguiente archivo M de MATLAB se puede desarrollar directamente, a partir del pseudocódigo dado en la figura 2.7. Escriba lo siguiente en el Editor/Debugger de MATLAB:

```
g=9.81;
m=input('mass (kg):');
cd=12.5;
ti=0;
tf=2;
vi=0;
dt=0.1;
t = ti;
v = vi;
h = dt;
while (1)
  if t + dt > tf
    h = tf – t;
  end
  dvdt = g – (cd / m) * v;
  v = v + dvdt * h;
  t = t + h;
  if t >= tf, break, end
end
disp('velocity (m/s):')
disp(v)
```

Guarde este archivo como numpara.m, vuelva al modo de comandos y córralo dando numpara. Obtendrá la siguiente salida:

```
masa (kg): 100

velocity (m/s):
   17.4559
```

Por último vamos a convertir este archivo M en una función. Esto se puede hacer en el siguiente archivo M basado en el pseudocódigo de la figura 2.7:

```
function yy = euler (dt,ti,tf,yi,m,cd)
t = ti;
y = yi;
h = dt;
while (1)
  if t + dt > tf
    h = tf – t;
  end
  dydt = dy(t, y, m, cd);
  y = y + dydt * h;
  t = t + h;
  if t >= tf, break, end
end
yy = y;
```

Guarde este archivo como euler.m y después cree otro archivo M para calcular la derivada,

```
function dydt = dy(t, v, m, cd)
g = 9.81;
dydt = g – (cd / m) * v;
```

Guarde este archivo como dy.m y regrese al modo de comandos. Para llamar la función y ver el resultado, teclee los siguientes comandos

a) **Pseudocódigo**	*b*) **MATLAB**

IF/THEN:

```
IF condición THEN
  Bloque verdadero
ENDIF
```

```
if b ~= 0
  r1 = -c / b;
end
```

IF/THEN/ELSE:

```
IF condición THEN
  Bloque verdadero
ELSE
  Bloque falso
ENDIF
```

```
if a < 0
  b = sqrt(abs(a));
else
  b = sqrt(a);
end
```

IF/THEN/ELSEIF:

```
IF condición1 THEN
  Bloque1
ELSEIF condición2
  Bloque2
ELSEIF condición3
  Bloque3
ELSE
  Bloque4
ENDIF
```

```
if class == 1
  x = x + 8;
elseif class < 1
  x = x - 8;
elseif class < 10
  x = x - 32;
else
  x = x - 64;
end
```

CASE:

```
SELECT CASE Expresión de prueba
  CASE Valor1
    Bloque1
  CASE Valor2
    Bloque2
  CASE Valor3
    Bloque3
  CASE ELSE
    Bloque4
END SELECT
```

```
switch a + b
  case 1
    x = -5;
  case 2
    x = -5 - (a + b) / 10;
  case 3
    x = (a + b) / 10;
  otherwise
    x = 5;
end
```

DOEXIT:

```
DO
  Bloque1
  IF condición EXIT
  Bloque2
ENDDO
```

```
while (1)
  i = i + 1;
  if i >= 10, break, end
  j = i*x;
end
```

LOOP CONTROLADO POR CONTADOR:

```
DOFOR i = inicio, fin, incremento
  Bloque
ENDDO
```

```
for i = 1:2:10
  x = x + i;
end
```

Figura 2.9
Estructuras de control fundamentales en *a*) pseudocódigo y *b*) lenguaje de programación en MATLAB.

```
>> m=68.1;
>> cd=12.5;
>> ti=0;
>> tf=2.;
>> vi=0;
>> dt=0.1;
>> euler(dt,ti,tf,vi,m,cd)
```

Una vez dado el último comando, se desplegará el resultado

```
ans =
16.5478
```

La combinación del ambiente de MATLAB con el lenguaje de programación para los archivos M nos abre un mundo de posibilidades para la solución de problemas en ingeniería. En el siguiente capítulo veremos cómo se hace esto.

2.6 MATHCAD

Mathcad intenta puentear la brecha entre las hojas de trabajo como Excel y los notepads. Originalmente fue desarrollado por Allen Razdow, del MIT, quien fuera cofundador de Mathsoft, Inc., y publicara la primera versión comercial en 1986. Actualmente, Mathsoft forma parte de Parametric Technology Corporation (PTC), y Mathcad se encuentra en la versión 15.

Mathcad es esencialmente un notepad interactivo que permite a los ingenieros y a los científicos realizar diversas tareas matemáticas, de manejo de datos y de tipo gráfico. La información y las ecuaciones se ingresan a un ambiente de diseño de "pizarra blanca" (*whiteboard*), sustancialmente similar a una hoja de papel. A diferencia de una herramienta de programación o de una hoja de trabajo, la interfaz de Mathcad acepta y muestra notación matemática natural, usando teclas o clics en las opciones del menú, sin necesidad de programación. Como las hojas de cálculo contienen cálculos en vivo, un solo golpe de teclado, que cambie un dato ingresado o una ecuación, devuelve instantáneamente un resultado actualizado.

Mathcad puede realizar tareas ya sea en modo numérico o en modo simbólico. En el primero, las funciones y los operadores de Mathcad dan respuestas numéricas, mientras que en el segundo, los resultados se dan como expresiones o ecuaciones generales. Maple V, un amplio paquete matemático simbólico, es la base del modo simbólico, y se incorporó a Mathcad en 1993.

Mathcad tiene una diversidad de funciones y operadores que permite la cómoda implementación de muchos de los métodos numéricos que se desarrollan en este libro. Esto se describirá a detalle en los capítulos siguientes. En el caso de que usted no esté familiarizado con Mathcad, el apéndice C proporciona también una introducción sobre el uso de este poderoso software.

2.7 OTROS LENGUAJES Y BIBLIOTECAS

En las secciones 2.4 y 2.5 mostramos cómo se escribe una función en Excel o MATLAB, para el método de Euler, a partir de un algoritmo expresado en pseudocódigo. Funciones semejantes se escriben en los lenguajes de alto nivel como Fortran 90 y C++. Por ejemplo, una función en Fortran 90 para el método de Euler es

```
   Function Euler(dt, ti, tf, yi, m, cd)

   REAL dt, ti, tf, yi, m, cd
   Real h, t, y, dydt

   t = ti
   y = yi
   h = dt
   Do
     If (t + dt > tf) Then
       h = tf – t
     End If
     dydt = dy(t, y, m, cd)
     y = y + dydt * h
     t = t + h
     If (t >= tf) Exit
End Do
Euler = y
End Function
```

En C el resultado sería bastante similar a la función escrita en MATLAB. El punto es que una vez que se ha desarrollado bien un algoritmo estructurado en pseudocódigo, es fácil implementarlo en diversos ambientes de programación.

En este libro daremos al lector procedimientos bien estructurados escritos en pseudocódigo. Esta colección de algoritmos constituirá una biblioteca numérica, que se puede usar para realizar tareas numéricas específicas con diversas herramientas de software y lenguajes de programación.

Además de tener sus propios programas, usted debe recordar que las bibliotecas comerciales de programación tienen muchos procedimientos numéricos útiles. Por ejemplo, la biblioteca *Numerical Recipe* contiene una gran variedad de algoritmos escritos en Fortran y C.[5] Estos procedimientos se describen tanto en libros (por ejemplo, Press *et al*., 2007) como en forma electrónica.

[5] Los procedimientos Numerical Recipe también están disponibles en libro y en formato electrónico para Pascal, MS BASIC y MATLAB. En http://www.nr.com se puede encontrar la información sobre todos los productos Numerical Recipe.

PROBLEMAS

2.1 Escriba el pseudocódigo para implementar el diagrama de flujo que se ilustra en la figura P2.1. Asegúrese de incluir la indentación apropiada para que la estructura sea clara.

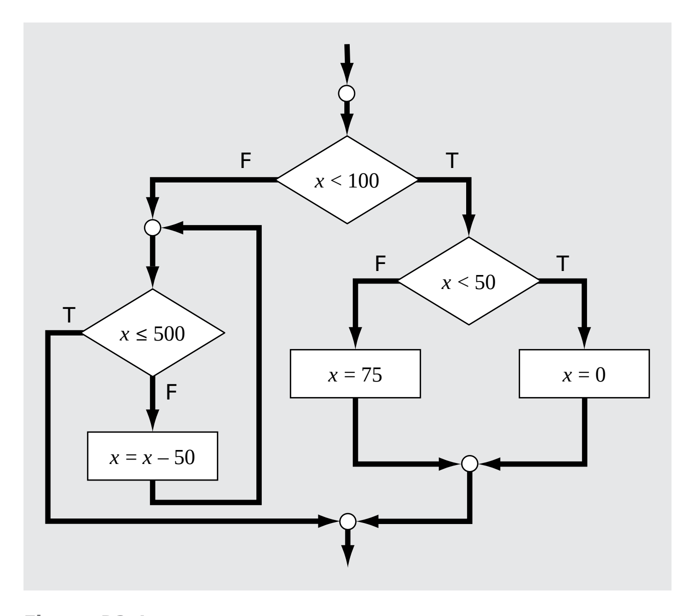


Figura P2.1

2.2 Vuelva a escribir el siguiente pseudocódigo usando las sangrías adecuadas:

```
DO
j = j + 1
x = x + 5
IF x > 5 THEN
y = x
ELSE
y = 0
ENDIF
z = x + y
IF z > 50 EXIT
ENDDO
```

2.3 Desarrolle, depure y documente un programa para determinar las raíces de una ecuación cuadrática, $ax^2 + bx + c$, en cualquier lenguaje de alto nivel, o de macros, de su elección. Utilice un procedimiento de subrutina para calcular las raíces (sean reales o complejas). Ejecute corridas de prueba para los casos en que *a*) $a = 1$, $b = 6$, $c = 2$; *b*) $a = 0$, $b = -4$, $c = 1.6$; *c*) $a = 3$, $b = 2.5$, $c = 7$.

2.4 La función seno puede evaluarse por medio de la serie infinita siguiente:

$$\text{sen}\, x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$$

Escriba un algoritmo para implementar esta fórmula de modo que calcule e imprima los valores de sen *x* conforme se agregue cada término de la serie. En otras palabras, calcule e imprima la secuencia de valores para

$$\text{sen}\, x = x$$

$$\text{sen}\, x = x - \frac{x^3}{3!}$$

$$\text{sen}\, x = x - \frac{x^3}{3!} + \frac{x^5}{5!}$$

hasta el término de orden *n* que usted elija. Para cada uno de los valores anteriores, calcule y haga que se muestre el error porcentual relativo:

$$\%\ \text{error} = \frac{\text{valor verdadero} - \text{aproximación con la serie}}{\text{valor verdadero}} \times 100\%$$

Escriba el algoritmo para *a*) diagrama de flujo estructurado, y *b*) pseudocódigo.

2.5 Desarrolle, depure y documente un programa para el problema 2.4 en cualquier lenguaje de alto nivel o de macros, de su elección. Emplee la función seno de la biblioteca de su computadora para determinar el valor verdadero. Haga que el programa imprima en cada paso la serie de aproximación y el error. Como caso de prueba, utilice el programa para calcular valores desde sen(1.5) hasta incluir el término $x^{15}/15!$ Interprete los resultados.

2.6 El algoritmo siguiente está diseñado para determinar la calificación de un curso que consiste en cuestionarios, tareas y un examen final:

Paso 1: Introducir la clave y nombre del curso.
Paso 2: Introducir factores de ponderación para los cuestionarios (C), tareas (T) y examen final (E).
Paso 3: Introducir las calificaciones de las preguntas y determinar su promedio (PC).
Paso 4: Introducir las calificaciones de las tareas y determinar su promedio (PT).
Paso 5: Si el curso tiene una calificación final, continuar con el paso 6. Si no, ir al paso 9.
Paso 6: Introducir la calificación del examen final (F).
Paso 7: Determinar la calificación promedio, CP, de acuerdo con

$$\text{CP} = \frac{\text{C} \times \text{PC} + \text{T} \times \text{PT} + \text{E} \times \text{F}}{\text{C} + \text{T} + \text{E}} \times 100\%$$

Paso 8: Ir al paso 10.
Paso 9: Determinar la calificación promedio, CP, de acuerdo con

$$\text{CP} = \frac{\text{C} \times \text{PC} + \text{T} \times \text{PT}}{\text{C} + \text{T}} \times 100\%$$

Paso 10: Imprimir la clave y nombre del curso, y la calificación promedio.
Paso 11: Finalizar el cálculo.

a) Escriba un pseudocódigo bien estructurado para implementar este algoritmo.
b) Escriba, depure y documente un programa estructurado de computadora basado en este algoritmo. Pruébelo con los datos siguientes para calcular una calificación sin el examen final, y otra con éste. C = 30; T = 40; E = 30; cuestionario = 98, 95, 90, 60 y 99; tareas = 98, 95, 86, 100, 100 y 77, y examen final = 91.

2.7 El método antiguo de *dividir y promediar*, para obtener el valor aproximado de la raíz cuadrada de cualquier número positivo *a* se puede formular como

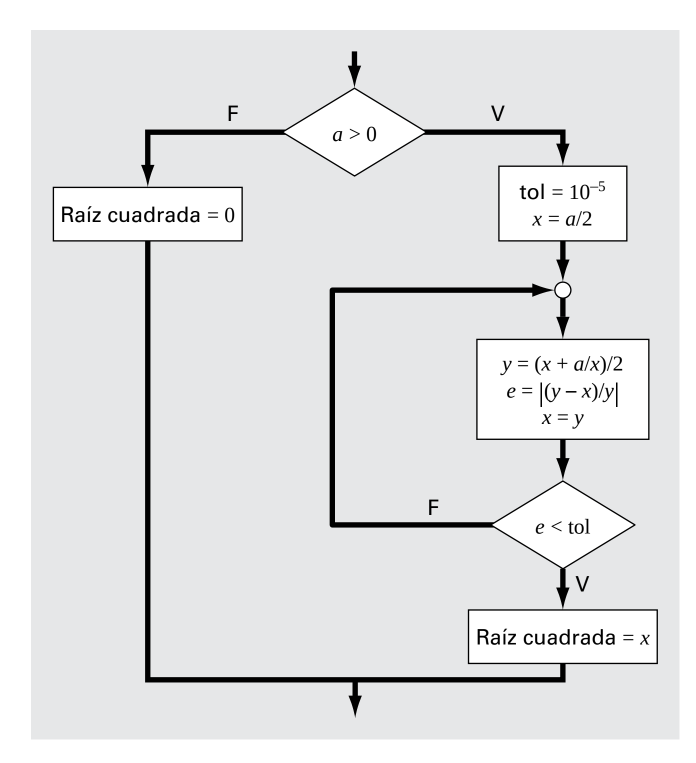


Figura P2.7

$$x = \frac{x + a/x}{2}$$

a) Escriba un pseudocódigo bien estructurado para implementar este algoritmo como se ilustra en la figura P2.7. Utilice la indentación apropiada para que la estructura sea clara.
b) Desarrolle, depure y documente un programa para implementar esta ecuación en cualquier lenguaje de alto nivel, o de macros, de su elección. Estructure su código de acuerdo con la figura P2.7.

2.8 Se invierte cierta cantidad de dinero *P* en una cuenta en la que el interés se capitaliza al final del periodo. Debe determinarse el valor futuro, *F*, que se obtiene con cierta tasa de interés, *i*, después de *n* periodos, por medio de la fórmula siguiente:

$$F = P\,(1 + i)^n$$

Escriba un programa que calcule el valor futuro de una inversión para cada año, desde 1 hasta *n*. La entrada para la función debe incluir la inversión inicial, *P*, la tasa de interés, *i* (en forma decimal), y el número de años, *n*, para el que ha de calcularse el valor futuro. La salida debe consistir en una tabla con encabezados y columnas para *n* y *F*. Corra el programa para *P* = $100 000, *i* = 0.04 y *n* = 11 años.

2.9 Las fórmulas económicas están disponibles para calcular los pagos anuales de préstamos. Suponga que obtiene en préstamo cierta cantidad de dinero *P* y acuerda devolverla en *n* pagos anuales con una tasa de interés de *i*. La fórmula para calcular el pago anual *A* es:

$$A = P\frac{i(1+i)^n}{(1+i)^n - 1}$$

Escriba un programa para calcular *A*. Pruébelo con *P* = $55 000 y una tasa de interés de 6.6% (*i* = 0.066). Calcule los resultados para *n* = 1, 2, 3, 4 y 5, y muestre los resultados en forma de tabla con encabezados y columnas para *n* y *A*.

2.10 La temperatura promedio diaria para cierta área se aproxima por medio de la función siguiente,

$$T = T_{media} + (T_{máxima} - T_{media}) \cos(\omega(t - t_{máxima}))$$

donde T_{media} = temperatura promedio anual, $t_{máxima}$ = temperatura máxima, ω = frecuencia de la variación anual (= $2\pi/365$) y $t_{máxima}$ = día de la temperatura máxima (≅ 205 d). Desarrolle un programa que calcule la temperatura promedio entre dos días del año para una ciudad en particular. Pruébelo para *a*) enero-febrero (*t* = 0 a 59) en Miami, Florida (T_{media} = 22.1°C; $T_{máxima}$ = 28.3°C), y *b*) julio-agosto (*t* = 180 a 242) en Boston, Massachusetts (T_{media} = 10.7°C; $T_{máxima}$ = 22.9°C).

2.11 Desarrolle, depure y pruebe un programa en cualquier lenguaje de alto nivel, o de macros, de su elección, a fin de calcular la velocidad del paracaídas que cae como se explicó en el ejemplo 1.2. Diseñe el programa de modo que permita al usuario introducir valores para el coeficiente de arrastre y la masa. Pruebe el programa con la reproducción de los resultados del ejemplo 1.2. Repita el cálculo pero utilice tamaños de paso de 1 y 0.5 s. Compare sus resultados con la solución analítica que se obtuvo previamente, en el ejemplo 1.1. Un tamaño de paso más pequeño, ¿hace que los resultados sean mejores o peores? Explique sus resultados.

2.12 El método de la burbuja es una técnica de ordenamiento ineficiente pero fácil de programar. La idea que subyace al ordenamiento consiste en avanzar hacia abajo a través de un arreglo, comparar los pares adyacentes e intercambiar los valores si no están en orden. Para

que este método ordene por completo un arreglo, es necesario que se repita muchas veces. Conforme se avanza para un ordenamiento en orden ascendente, los elementos más pequeños del arreglo parecen ascender como burbujas. Eventualmente, habrá un paso por el arreglo que ya no requiera intercambios. En ese momento, el arreglo estará ordenado. Después del primer paso, el valor más grande cae directamente hasta el fondo. En consecuencia, el segundo paso sólo tiene que proceder del segundo al último valor, y así sucesivamente. Desarrolle un programa que tome un arreglo de 20 números al azar y los ordene en forma ascendente con la técnica de la burbuja (véase la figura P2.12).

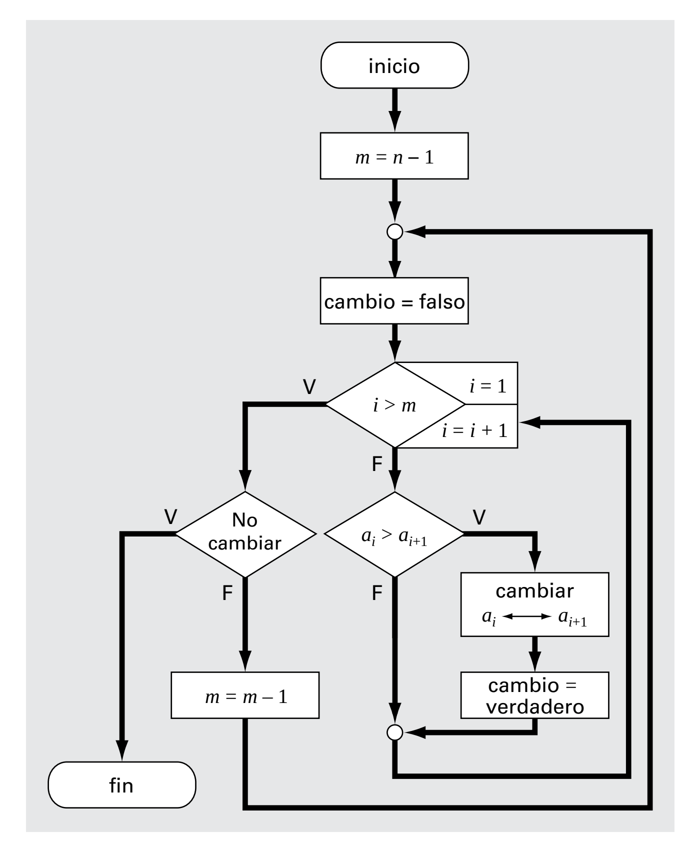


Figura P2.12

2.13 En la figura P2.13 se muestra un tanque cilíndrico con base cónica. Si el nivel del líquido está muy bajo en la parte cónica, el volumen simplemente es el volumen del cono de líquido. Si el nivel del líquido está en la parte cilíndrica, el volumen total de líquido incluye la parte cónica llena y la parte cilíndrica parcialmente llena. Escriba un procedimiento bien estructurado de función para calcular el volumen del tanque como función de los valores dados de R y d. Utilice estructuras de control de decisiones (como If/Then, ElseIf, Else, End If). Diseñe la función de modo que produzca el volumen en todos los casos en los que la profundidad sea menor que $3R$. Genere un mensaje de error (“Sobrepasado”) si se rebasa la altura del tanque, es decir, $d > 3R$. Pruébelo con los datos siguientes:

R	1	1	1	1
d	0.5	1.2	3.0	3.1

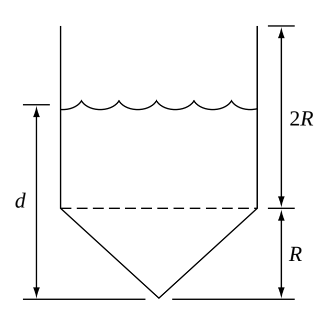


Figura P2.13

2.14 Se requieren dos distancias para especificar la ubicación de un punto en relación con el origen en un espacio de dos dimensiones (véase la figura P2.14):

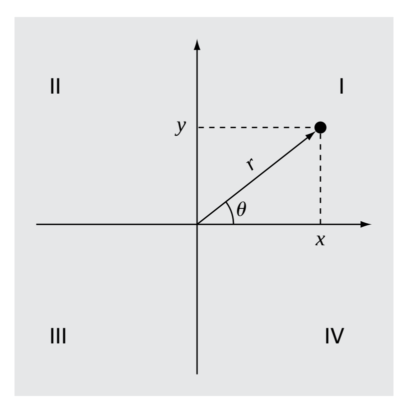


Figura P2.14

- Las distancias horizontal y vertical (x, y) en coordenadas cartesianas.
- El radio y el ángulo (r, θ) en coordenadas radiales.

Es relativamente fácil calcular las coordenadas cartesianas (x, y) sobre la base de las coordenadas polares (r, θ). El proceso inverso no es tan simple. El radio se calcula con la fórmula que sigue:

$$r = \sqrt{x^2 + y^2}$$

Si las coordenadas quedan dentro del primer o cuarto cuadrante (p. ej., $x > 0$), entonces se emplea una fórmula sencilla para el cálculo de θ:

$$\theta = \tan^{-1}\left(\frac{y}{x}\right)$$

La dificultad surge en los demás casos. La tabla siguiente resume las posibilidades:

x	*y*	*θ*
<0	>0	$\tan^{-1}(y/x) + \pi$
<0	<0	$\tan^{-1}(y/x) - \pi$
<0	=0	π
=0	>0	$\pi/2$
=0	<0	$-\pi/2$
=0	=0	0

a) Escriba un diagrama de flujo bien estructurado para un procedimiento de subrutina a fin de calcular r y θ como función de x y y. Exprese los resultados finales para θ, en grados.

b) Escriba un procedimiento bien estructurado de función con base en el diagrama de flujo. Pruebe el programa de modo que se llene la tabla que sigue:

x	*y*	*r*	*θ*
1	0		
1	1		
0	1		
–1	1		
–1	0		
–1	–1		
0	–1		
1	–1		
0	0		

2.15 Desarrolle un procedimiento bien estructurado de función que lea una calificación numérica entre 0 y 100 y devuelva una letra, de acuerdo con el esquema siguiente:

Letra	Criterio
A	90 ≤ calificación numérica ≤ 100
B	80 ≤ calificación numérica < 90
C	70 ≤ calificación numérica < 80
D	60 ≤ calificación numérica < 70
F	calificación numérica < 60

2.16 Desarrolle un procedimiento bien estructurado de función para determinar *a*) el factorial de un número; *b*) el valor más pequeño de un vector, y *c*) el promedio de los valores de un vector.

2.17 Desarrolle programas bien estructurados para *a*) determinar la raíz cuadrada de la suma de los cuadrados de los elementos de un arreglo bidimensional (p. ej., una matriz), y *b*) normalizar una matriz por medio de dividir cada renglón entre el valor absoluto máximo en el renglón de modo que el elemento mayor en cada renglón sea 1.

2.18 Las *funciones segmentadas* son algunas veces útiles cuando la relación entre una variable dependiente y una independiente no se puede representar adecuadamente con una sola ecuación. Por ejemplo, la velocidad de un cohete se puede describir mediante

$$v(t) = \begin{cases} 11t^2 - 5t & 0 \le t \le 10 \\ 1100 - 5t & 10 \le t \le 20 \\ 50t + 2(t-20)^2 & 20 \le t \le 30 \\ 1520e^{-0.2(t-30)} & t > 30 \\ 0 & \text{de otro modo} \end{cases}$$

Desarrolle una función bien estructurada para calcular v como función de t. Luego use esa función para generar una tabla de v contra t para t = –5 a 50 a incrementos de 0.5.

2.19 Desarrolle una función bien estructurada para determinar los días transcurridos en un año. La función debe pasar por tres valores: `mo` = el mes (1-12), `da` = el día (1-31), y `leap` (bisiesto) (0 para año no bisiesto, y 1 para año bisiesto). Pruébela para 1 de enero de 1999; 29 de febrero de 2000; 1 de marzo de 2001; 21 de junio de 2002, y 31 de diciembre de 2004. *Sugerencia:* Una buena manera de hacer esto combina las estructuras `for` y `switch`.

2.20 Desarrolle una función bien estructurada para determinar los días transcurridos en un año. La primera línea de una función se debe establecer como

```
function nd = days(mo, da, year)
```

donde `mo` = el mes (1-12), `da` = el día (1-31) y `year` = año. Pruébelo para el 1 de enero de 1999; 29 de febrero de 2000; 1 de marzo de 2001; 21 de junio de 2002, y 31 de diciembre de 2004.

2.21 Las ecuaciones de Manning se pueden usar para calcular la velocidad del agua en un canal abierto de sección rectangular,

$$U = \frac{\sqrt{S}}{n}\left(\frac{BH}{B+2H}\right)^{2/3}$$

donde U = velocidad (m/s), S = pendiente del canal, n = coeficiente de rugosidad, B = anchura (m) y H = profundidad (m). Los siguientes datos están disponibles para cinco canales:

n	*S*	*B*	*H*
0.035	0.0001	10	2
0.020	0.0002	8	1
0.015	0.0010	20	1.5
0.030	0.0007	24	3
0.022	0.0003	15	2.5

Escriba un programa bien estructurado que calcule la velocidad de cada uno de estos canales. Haga que el programa muestre los datos de entrada junto con la velocidad completada en forma tabular, donde la velocidad es la quinta columna. Incluya encabezados para las columnas.

2.22 La figura P2.22 muestra una viga con un solo apoyo. Con el uso de funciones de singularidad, el desplazamiento a lo largo de la viga se puede expresar por la ecuación:

$$u_y(x) = \frac{-5}{6}[\langle x-0\rangle^4 - \langle x-5\rangle^4] + \frac{15}{6}\langle x-8\rangle^3 + 75\langle x-7\rangle^2 + \frac{57}{6}x^3 - 238.25x$$

Por definición, la función de singularidad se puede expresar como sigue:

$$\langle x-a\rangle^n = \begin{Bmatrix} (x-a)^n & \text{cuando } x > a \\ 0 & \text{cuando } x \le a \end{Bmatrix}$$

Desarrolle un programa que cree una gráfica de desplazamiento contra distancia a lo largo de la viga, x. Observe que x = 0 al extremo izquierdo de la viga.

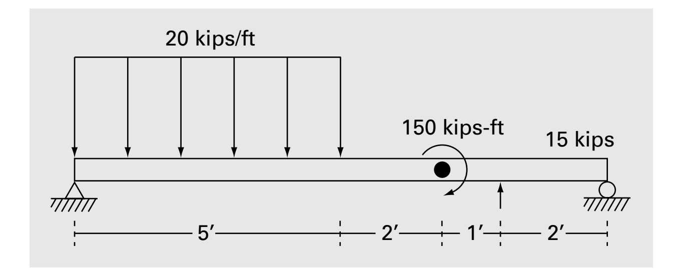


Figura P2.22

2.23 El volumen V de líquido en un cilindro hueco horizontal de radio r y longitud L se relaciona con la profundidad del líquido h por

$$V = \left[r^2 \cos^{-1}\left(\frac{r-h}{r}\right) - (r-h)\sqrt{2rh - h^2} \right] L$$

Desarrolle una función bien estructurada para crear una gráfica de volumen contra profundidad. Pruebe el programa para $r = 2$ m y $L = 5$ m.

2.24 Desarrolle un programa bien estructurado para calcular la velocidad de un paracaidista como función del tiempo, usando el método de Euler. Pruebe su programa para el caso en el que $m = 80$ kg y $c = 10$ kg/s. Realice el cálculo a partir de $t = 0$ a 20 s, con un tamaño de paso de 2 s. Use una condición inicial de que el paracaidista tiene una velocidad ascendente de 20 m/s a $t = 0$. A $t = 10$ s, suponga que el paracaídas se abre instantáneamente de modo que el coeficiente de arrastre salta a 50 kg/s.

2.25 El pseudocódigo en la figura P2.25 calcula el factorial. Exprese este algoritmo como una función bien estructurada en el lenguaje que elija. Pruébelo calculando 0! y 5! Además, pruebe la trampa de errores tratando de evaluar –2!

```
FUNCTION fac(n)
IF n ≥ 0 THEN
  x = 1
  DOFOR i = 1, n
    x = x · i
  END DO
  fac = x
ELSE
  display error message
  terminate
ENDIF
END fac
```

Figura P2.25

2.26 La altura de un pequeño cohete, y, se puede calcular como función del tiempo después del lanzamiento con la siguiente función segmentada:

$$y = 38.1454t + 0.13743t^3 \qquad 0 \le t < 15$$

$$y = 1\,036 + 130.909\,(t-15) + 6.18425\,(t-15)^2 - 0.428\,(t-15)^3 \qquad 15 \le t < 33$$

$$y = 2\,900 - 62.468\,(t-33) - 16.9274\,(t-33)^2 + 0.41796\,(t-33)^3 \qquad t > 33$$

Desarrolle una función bien estructurada de pseudocódigo para calcular y como función de t. Observe que si el usuario ingresa un valor negativo de t, o si el cohete ha tocado tierra ($y \le 0$), entonces devolver un valor de 0 para y. Asimismo, la función se debe invocar en el programa de llamado como `height (t)`. Escriba el algoritmo como *a*) pseudocódigo, o *b*) en el lenguaje de alto nivel que usted elija.

2.27 Como se describe en la figura P2.27, un tanque de agua consiste en un cilindro que tiene en la parte superior el frústum de un cono. Desarrolle una función bien estructurada en el lenguaje de alto nivel o macrolenguaje que usted elija para calcular el volumen dado el nivel de agua h (m) sobre el fondo del tanque. Diseñe la función de tal manera que devuelva un valor de cero para las h negativas, y el valor del volumen máximo lleno para las h mayores que la profundidad máxima del tanque. Dados los siguientes parámetros, $H_1 = 10$ m, $r_1 = 4$ m, $H_2 = 5$ m y $r_2 = 6.5$ m, pruebe su función usándola para calcular los volúmenes, y genere una gráfica como función del nivel, desde $h = -1$ hasta 16 m.

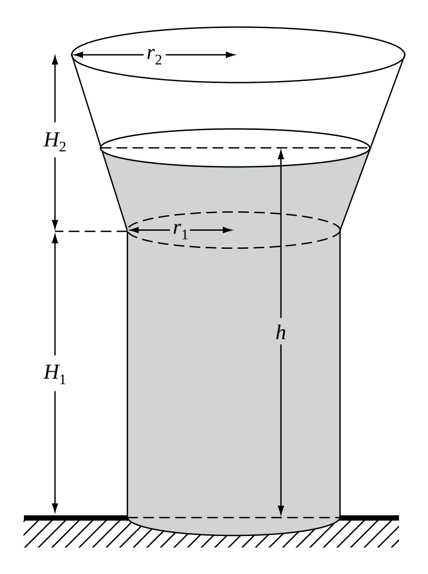


Figura P2.27

CAPÍTULO 3

Aproximaciones y errores de redondeo

A causa de que la mayor parte de los métodos expuestos en este libro son muy sencillos en su descripción y en sus aplicaciones, en este momento resulta tentador ir directamente al cuerpo principal del texto y averiguar el empleo de dichas técnicas. Sin embargo, entender el concepto de error es tan importante para utilizar en forma efectiva los métodos numéricos que los dos siguientes capítulos se eligieron para tratar el tema.

La importancia de los errores se mencionó por primera vez en el análisis de la caída del paracaidista en el capítulo 1. Recuerde que la velocidad de caída del paracaidista se determinó por métodos analíticos y numéricos. Aunque con la técnica numérica se obtuvo una aproximación a la solución analítica exacta, hubo cierta discrepancia o *error*, debido a que los métodos numéricos dan sólo una aproximación. En realidad fuimos afortunados en este caso porque teníamos la solución analítica que nos permitía calcular el error en forma exacta. Pero en muchos problemas de aplicación en ingeniería no es posible obtener la solución analítica; por lo tanto, no se pueden calcular con exactitud los errores en nuestros métodos numéricos. En tales casos debemos usar aproximaciones o estimaciones de los errores.

Tales errores son característicos de la mayoría de las técnicas descritas en este libro. En primera instancia, esto puede parecer contradictorio, ya que no coincide con la imagen que se tiene de una buena ingeniería. Los estudiantes y los practicantes de la ingeniería trabajan constantemente para limitar este tipo de errores en sus actividades. Cuando hacen un examen o realizan sus tareas, son sancionados, mas no premiados por sus errores. En la práctica profesional, los errores llegan a resultar costosos y, en algunas ocasiones, catastróficos. Si una estructura o un dispositivo falla, esto puede costar vidas.

Aunque la perfección es una meta digna de alabarse, es difícil, si no imposible, alcanzarla. Por ejemplo, a pesar de que el modelo obtenido mediante la segunda ley de Newton es una aproximación excelente, en la práctica jamás predecirá con exactitud la caída del paracaidista. Fenómenos como la velocidad del viento y alguna ligera variación de la resistencia del aire desviarían la predicción. Si tales desviaciones son sistemáticamente grandes o pequeñas, habría entonces que formular un nuevo modelo. No obstante, si su distribución es aleatoria y se agrupan muy cerca de la predicción, entonces las desviaciones se considerarían insignificantes y el modelo parecerá adecuado. Las aproximaciones numéricas también presentan discrepancias similares en el análisis. De nuevo, las preguntas son: ¿qué tanto error se presenta en los cálculos?, y ¿es tolerable?

Este capítulo y el siguiente cubren aspectos básicos relacionados con la identificación, cuantificación y minimización de dichos errores. En las primeras secciones se revisa la información referente a la cuantificación de los errores. En seguida, se estudia uno de los dos errores numéricos más comunes: errores de redondeo. Los *errores de redondeo* se deben a que la computadora tan sólo representa cantidades con un número finito de dígitos. En el siguiente capítulo nos ocuparemos de otra clase importante de error: el de truncamiento. El *error de truncamiento* es la discrepancia que se introduce por el hecho de que los métodos numéricos pueden emplear aproximaciones para representar operaciones y cantidades matemáticas exactas. Por último, se analizan los errores que no están relacionados directamente con el método numérico en sí. Éstos son equivocaciones, errores de formulación o del modelo, y la incertidumbre en la obtención de los datos.

3.1 CIFRAS SIGNIFICATIVAS

En esta obra se trata de manera extensa con aproximaciones que se relacionan con el manejo de números. En consecuencia, antes de analizar los errores asociados con los métodos numéricos, es útil repasar algunos conceptos básicos referentes a la representación aproximada de los números mismos.

Cuando se emplea un número para realizar un cálculo, debe haber seguridad de que pueda usarse con confianza. Por ejemplo, la figura 3.1 muestra un velocímetro y un odómetro (contador de kilometraje) de un automóvil. Con un simple vistazo al velocímetro se observa que el vehículo viaja a una velocidad comprendida entre 48 y 49 km/h. Como la aguja está más allá de la mitad entre las marcas del indicador, es posible asegurar que el automóvil viaja aproximadamente a 49 km/h. Tenemos confianza en este resultado, ya que dos o más individuos que hicieran esta lectura llegarían a la misma conclusión. Sin embargo, supongamos que se desea obtener una cifra decimal en la estimación de la velocidad. En tal caso, alguien podría decir 48.8, mientras que otra persona podría decir 48.9 km/h. Por lo tanto, debido a los límites del instrumento, únicamente se emplean con confianza los dos primeros dígitos. Para estimaciones del tercer dígito (o más allá) sólo se considerarían aproximaciones. Sería ridículo afirmar, considerando el velocímetro de la figura, que el automóvil viaja a 48.8642138 km/h. En contraste, el odómetro muestra hasta seis dígitos confiables. De la figura 3.1 se concluye que el automóvil ha recorrido un poco menos de 87 324.5 km durante su uso. Aquí el séptimo dígito (y los siguientes) resultan inciertos.

El concepto de cifras o dígitos significativos se ha desarrollado para designar formalmente la confiabilidad de un valor numérico. Las *cifras significativas* de un número son aquellas que pueden utilizarse en forma confiable. Se trata del número de dígitos que se ofrecen con certeza, más uno estimado. Por ejemplo, el velocímetro y el odómetro de la figura 3.1 muestran lecturas de hasta tres y siete cifras significativas, respectivamente. Para el velocímetro, los dos dígitos seguros son 48. Por convención se ajusta el dígito estimado a la mitad de la división de la escala menor en el instrumento de medición. Así, la lectura del velocímetro consistirá de las tres cifras significativas: 48.5. En forma similar, el odómetro dará una lectura con siete cifras significativas, 87 324.45.

Aunque, por lo común, determinar las cifras significativas de un número es un procedimiento sencillo, en algunos casos genera cierta confusión. Por ejemplo, los ceros no siempre son cifras significativas, ya que pueden usarse sólo para ubicar el punto decimal: los números 0.00001845, 0.0001845 y 0.001845 tienen cuatro cifras significativas. Asimismo, cuando se incluye ceros en números muy grandes, no queda claro cuántos son significativos. Por ejemplo, el número 45 300 puede tener tres, cuatro o cinco dígitos significativos, dependiendo de si los ceros se conocen o no con exactitud. La incertidumbre se puede eliminar utilizando la notación científica, donde 4.53×10^4, 4.530×10^4, 4.5300×10^4 muestran, respectivamente, que el número tiene tres, cuatro y cinco cifras significativas.

El concepto de cifras significativas tiene dos implicaciones importantes en el estudio de los métodos numéricos.

1. Como se mencionó en el problema de la caída del paracaidista, los métodos numéricos dan resultados aproximados. Por lo tanto, se deben desarrollar criterios para especificar qué tan confiables son dichos resultados. Una manera de hacerlo es en términos de cifras significativas. Por ejemplo, es posible afirmar que la aproximación es aceptable siempre y cuando sea correcta con cuatro cifras significativas.

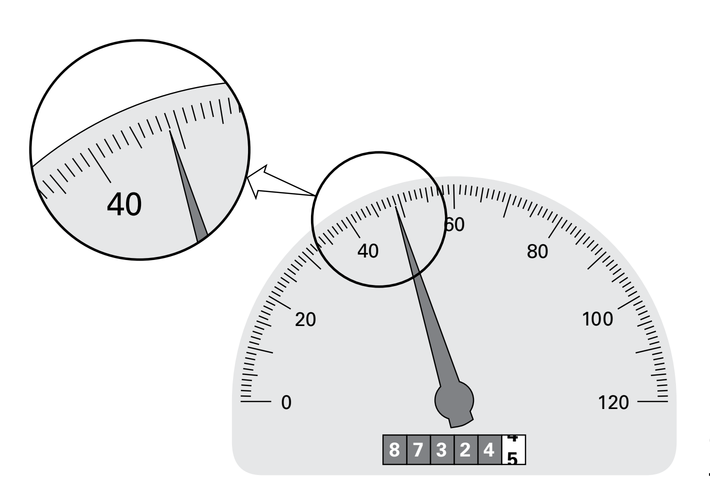


Figura 3.1
El velocímetro y el odómetro de un automóvil ejemplifican el concepto de cifras significativas.

2. Aunque ciertas cantidades como π, e o $\sqrt{7}$ representan cantidades específicas, no se pueden expresar exactamente con un número finito de dígitos. Por ejemplo,

   π = 3.141592653589793238462643...

   hasta el infinito. Como las computadoras retienen sólo un número finito de cifras significativas, tales números jamás se podrán representar con exactitud. A la omisión del resto de cifras significativas se le conoce como *error de redondeo*.

Los errores de redondeo y el uso de cifras significativas para expresar nuestra confianza en un resultado numérico se estudiarán con mayor detalle en las siguientes secciones. Además, el concepto de cifras significativas tendrá mucha importancia en la definición de exactitud y de precisión en la siguiente sección.

3.2 EXACTITUD Y PRECISIÓN

Los errores asociados con cálculos y medidas se pueden caracterizar con respecto a su exactitud y su precisión. La *exactitud* se refiere a qué tan cercano está el valor calculado o medido del valor verdadero. La *precisión* se refiere a qué tan cercanos se encuentran, unos de otros, diversos valores calculados o medidos.

Estos conceptos se ilustran gráficamente utilizando la analogía con una diana en la práctica de tiro. Los agujeros en cada blanco de la figura 3.2 se consideran como las predicciones con una técnica numérica; mientras que el centro del blanco representa la verdad. La *inexactitud* (conocida también como *sesgo*) se define como una desviación sistemática del valor verdadero. Por lo tanto, aunque los disparos en la figura 3.2*c* están más juntos que los de la figura 3.2*a*, los dos casos son igualmente inexactos, ya que ambos se centran en la esquina superior izquierda del blanco. La *imprecisión* (también llamada *incertidumbre*), por otro lado, se refiere a la magnitud en la dispersión de los disparos. Por consiguiente, aunque las figuras 3.2*b* y 3.2*d* son igualmente exactas (esto es, igualmente centradas respecto al blanco), la última es más precisa, pues los disparos están agrupados en forma más compacta.

Los métodos numéricos deben ser lo suficientemente exactos o sin sesgo para satisfacer los requisitos de un problema particular de ingeniería. También deben ser suficientemente precisos para ser adecuados en el diseño de la ingeniería. En este libro se usa el término *error* para representar tanto la inexactitud como la imprecisión en las predicciones. Con dichos conceptos como antecedentes, ahora analizaremos los factores que contribuyen al error en los cálculos numéricos.

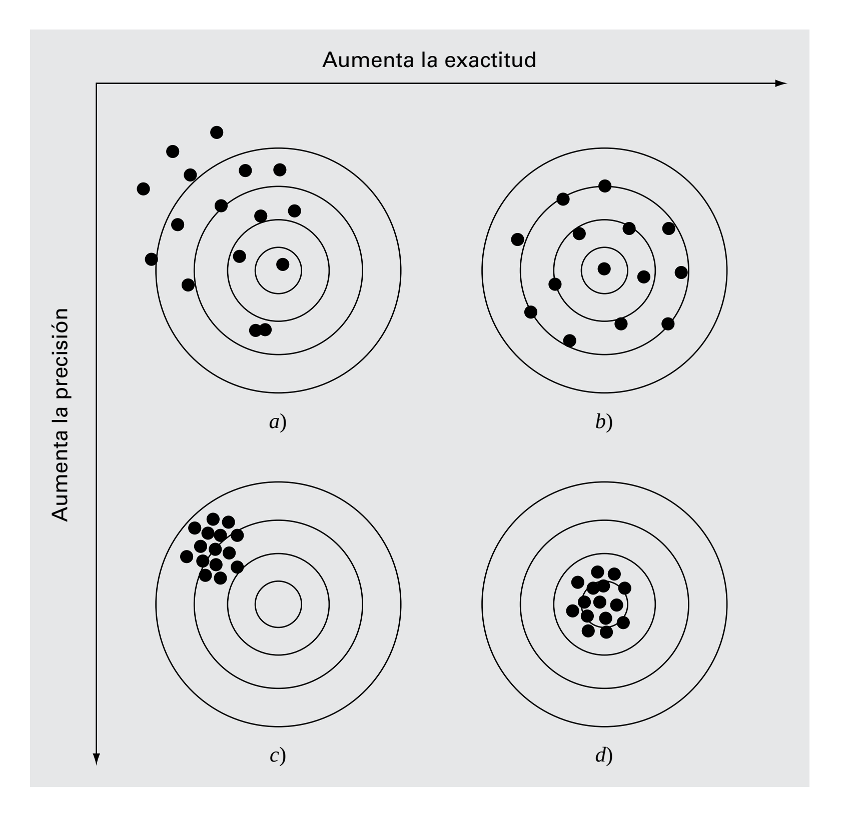


Figura 3.2
Un ejemplo de puntería ilustra los conceptos de exactitud y precisión. *a*) Inexacto e impreciso; *b*) exacto e impreciso; *c*) inexacto y preciso; *d*) exacto y preciso.

3.3 DEFINICIONES DE ERROR

Los errores numéricos surgen del uso de aproximaciones para representar operaciones y cantidades matemáticas exactas. Éstas incluyen los *errores de truncamiento* que resultan del empleo de aproximaciones como un procedimiento matemático exacto, y los *errores de redondeo* que se

producen cuando se usan números que tienen un límite de cifras significativas para representar números exactos. Para ambos tipos de errores, la relación entre el resultado exacto, o verdadero, y el aproximado está dada por

$$\text{Valor verdadero} = \text{valor aproximado} + \text{error} \tag{3.1}$$

Reordenando la ecuación (3.1) se encuentra que el error numérico es igual a la diferencia entre el valor verdadero y el valor aproximado, es decir

$$E_t = \text{valor verdadero} - \text{valor aproximado} \tag{3.2}$$

donde E_t se usa para denotar el valor exacto del error. El subíndice t indica que se trata del error "verdadero" (*true*). Como ya se mencionó brevemente, esto contrasta con los otros casos, donde se debe emplear una estimación "aproximada" del error.

Una desventaja en esta definición es que no toma en consideración el orden de la magnitud del valor que se estima. Por ejemplo, un error de un centímetro es mucho más significativo si se está midiendo un remache en lugar de un puente. Una manera de tomar en cuenta las magnitudes de las cantidades que se evalúan consiste en normalizar el error respecto al valor verdadero, es decir

$$\text{Error relativo fraccional verdadero} = \frac{\text{error verdadero}}{\text{valor verdadero}}$$

donde, como ya se mencionó en la ecuación (3.2), error = valor verdadero – valor aproximado. El error relativo también se puede multiplicar por 100% para expresarlo como

$$\varepsilon_t = \frac{\text{error verdadero}}{\text{valor verdadero}}\,100\% \tag{3.3}$$

donde ε_t denota el error relativo porcentual verdadero.

EJEMPLO 3.1 Cálculo de errores

Planteamiento del problema Suponga que se tiene que medir la longitud de un puente y la de un remache, y se obtiene 9 999 y 9 cm, respectivamente. Si los valores verdaderos son 10 000 y 10 cm, calcule *a*) el error verdadero y *b*) el error relativo porcentual verdadero en cada caso.

Solución

a) El error en la medición del puente es [ecuación (3.2)]

$$E_t = 10\,000 - 9\,999 = 1 \text{ cm}$$

y en la del remache es de

$$E_t = 10 - 9 = 1 \text{ cm}$$

b) El error relativo porcentual para el puente es [ecuación (3.3)]

$$\varepsilon_t = \frac{1}{10\,000}\,100\% = 0.01\%$$

y para el remache es de

$$\varepsilon_t = \frac{1}{10}\,100\% = 10\%$$

Por lo tanto, aunque ambas medidas tienen un error de 1 cm, el error relativo porcentual del remache es mucho mayor. Se concluye entonces que se ha hecho un buen trabajo en la medición del puente; mientras que la estimación para el remache dejó mucho que desear.

Observe que en las ecuaciones (3.2) y (3.3), E y ε tienen un subíndice t que significa que el error ha sido normalizado al valor verdadero. En el ejemplo 3.1 nos proporcionan este valor. Sin embargo,

en las situaciones reales a veces es difícil contar con tal información. En los métodos numéricos, el valor verdadero sólo se conocerá cuando se tengan funciones que se resuelvan analíticamente. Éste comúnmente será el caso cuando se estudie el comportamiento teórico de una técnica específica para sistemas simples. Sin embargo, en muchas aplicaciones reales, no se conoce *a priori* la respuesta verdadera. Entonces en dichos casos, una alternativa es normalizar el error, usando la mejor estimación posible al valor verdadero; es decir, para la aproximación misma, como en

$$\varepsilon_a = \frac{\text{error aproximado}}{\text{valor aproximado}} 100\% \tag{3.4}$$

donde el subíndice *a* significa que el error está normalizado a un valor aproximado. Observe también que en aplicaciones reales la ecuación (3.2) no se puede usar para calcular el término del error de la ecuación (3.4). Uno de los retos que enfrentan los métodos numéricos es el de determinar estimaciones del error en ausencia del conocimiento de los valores verdaderos. Por ejemplo, ciertos métodos numéricos usan un *método iterativo* para calcular los resultados. En tales métodos se hace una aproximación considerando la aproximación anterior. Este proceso se efectúa varias veces, o de forma iterativa, para calcular en forma sucesiva, esperando cada vez mejores aproximaciones. En tales casos, el error a menudo se calcula como la diferencia entre la aproximación previa y la actual. Por lo tanto, el error relativo porcentual está dado por

$$\varepsilon_a = \frac{\text{aproximación actual} - \text{aproximación anterior}}{\text{aproximación actual}} 100\% \tag{3.5}$$

En capítulos posteriores se explicarán con detalle éste y otros métodos para expresar errores.

Los signos de las ecuaciones (3.2) a (3.5) pueden ser positivos o negativos. Si la aproximación es mayor que el valor verdadero (o la aproximación previa es mayor que la aproximación actual), el error es negativo; si la aproximación es menor que el valor verdadero, el error es positivo. También en las ecuaciones (3.3) a (3.5), el denominador puede ser menor a cero, lo cual también llevaría a un error negativo. A menudo, cuando se realizan cálculos, no importa mucho el signo del error, sino más bien que su valor absoluto porcentual sea menor que una tolerancia porcentual prefijada ε_s. Por lo tanto, es útil emplear el valor absoluto de las ecuaciones (3.2) a (3.5). En tales casos, los cálculos se repiten hasta que

$$|\varepsilon_a| < \varepsilon_s \tag{3.6}$$

Si se cumple la relación anterior, entonces se considera que el resultado obtenido está dentro del nivel aceptable fijado previamente ε_s. Observe que en el resto del texto en general emplearemos exclusivamente valores absolutos cuando utilicemos errores relativos.

Es conveniente también relacionar estos errores con el número de cifras significativas en la aproximación. Es posible demostrar (Scarborough, 1966) que si el siguiente criterio se cumple, se tendrá la seguridad que el resultado es correcto en *al menos n* cifras significativas.

$$\varepsilon_s = (0.5 \times 10^{2-n})\% \tag{3.7}$$

EJEMPLO 3.2 Estimación del error con métodos iterativos

Planteamiento del problema En matemáticas con frecuencia las funciones se representan mediante series infinitas. Por ejemplo, la función exponencial se calcula usando

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} \tag{E3.2.1}$$

Así cuanto más términos se le agreguen a la serie, la aproximación será cada vez más una mejor estimación del valor verdadero de e^x. La ecuación (E3.2.1) se conoce como *expansión en series de Maclaurin*.

Empezando con el primer término $e^x = 1$ y agregando término por término, estime el valor de $e^{0.5}$. Después de agregar cada término, calcule el error relativo porcentual verdadero y aproximado usando las ecuaciones (3.3) y (3.5), respectivamente. Observe que el valor verdadero es $e^{0.5} = 1.648721\ldots$

Agregue términos hasta que el valor absoluto del error aproximado ε_a sea menor que un criterio de error preestablecido ε_s con tres cifras significativas.

Solución En primer lugar la ecuación (3.7) se emplea para determinar el criterio de error que asegura que un resultado sea correcto en al menos tres cifras significativas:

$$\varepsilon_s = (0.5 \times 10^{2-3})\% = 0.05\%$$

Por lo tanto, se agregarán términos a la serie hasta que ε_a sea menor que este valor.

La primera estimación es igual a la ecuación (E3.2.1) con un solo término. Entonces, la primera estimación es igual a 1. La segunda estimación se obtiene agregando el segundo término, así:

$$e^x = 1 + x$$

y para $x = 0.5$,

$$e^{0.5} = 1 + 0.5 = 1.5$$

Esto representa el error relativo porcentual verdadero de [ecuación (3.3)]

$$\varepsilon_t = \frac{1.648721 - 1.5}{1.648721} 100\% = 9.02\%$$

La ecuación (3.5) se utiliza para determinar una estimación aproximada del error, dada por:

$$\varepsilon_a = \frac{1.5 - 1}{1.5} 100\% = 33.3\%$$

Como ε_a no es menor que el valor requerido ε_s, se deben continuar los cálculos agregando otro término, $x^2/2!$, repitiendo el cálculo del error. El proceso continúa hasta que $\varepsilon_a < \varepsilon_s$. Todos los cálculos se resumen de la siguiente manera:

Términos	Resultado	ε_t (%)	ε_a (%)
1	1	39.3	
2	1.5	9.02	33.3
3	1.625	1.44	7.69
4	1.645833333	0.175	1.27
5	1.648437500	0.0172	0.158
6	1.648697917	0.00142	0.0158

Así, después de usar seis términos, el error aproximado es menor que $\varepsilon_s = 0.05\%$, y el cálculo termina. Sin embargo, observe que, ¡el resultado es exacto con cinco cifras significativas! en vez de tres cifras significativas. Esto se debe a que, en este caso, las ecuaciones (3.5) y (3.7) son conservadoras. Es decir, aseguran que el resultado es, por lo menos, tan bueno como lo especifican. Aunque, como se analiza en el capítulo 6, éste no es siempre el caso al usar la ecuación (3.5), que es verdadera en la mayoría de las veces.

3.3.1 Algoritmo de computadora para cálculos iterativos

Muchos de los métodos numéricos que se describen en el resto de este libro utilizan cálculos iterativos del tipo que se ilustra en el ejemplo 3.2. Todos ellos conllevan la resolución de un problema matemático mediante el cálculo de aproximaciones sucesivas a la solución, comenzando por una conjetura inicial.

La implementación en computadora de tales soluciones iterativas incluye ciclos de computación. Como se vio en la sección 2.1.1, éstos vienen en dos formas básicas: ciclos controlados por cuenta y *loops* de decisión. La mayor parte de las soluciones iterativas usan ciclos de decisión. Así, en vez de emplear un número preespecificado de iteraciones, el proceso típicamente se repite hasta que un estimado aproximado de error cae por debajo de un criterio de detención, como en el ejemplo 3.2.

En la figura 3.3 se presenta un pseudocódigo para un cálculo iterativo genérico. A la función se le pasa un valor (`val`) junto con criterio de error de detención (`es`), y un valor máximo permisible

de iteraciones (`maxit`). El valor típicamente es: 1) un valor inicial, o 2) el valor para el cual se debe hacer el cálculo iterativo.

La función inicializa primero tres variables, que incluyen: 1) una variable `iter` que lleva la cuenta del número de iteraciones, 2) una variable `sol` que contiene el estimado actual de la solución y 3) una variable `ea` que contiene el porcentaje aproximado de error relativo. Observe que `ea` al inicio se ajusta a un valor de 100 para asegurar que el bucle se ejecute por lo menos una vez.

Después de estas inicializaciones sigue el bucle de decisión que realmente implementa el cálculo iterativo. Antes de generar una nueva solución, `sol` se asigna primero a `solold`. Luego se calcula un nuevo valor de `sol`, y se incrementa el contador de iteraciones. Si el nuevo valor de `sol` es diferente de cero, se determina el porcentaje de error relativo `ea`. En seguida se prueban los criterios de detención. Si ambos son falsos, se repite el bucle. Si cualquiera de ellos es verdadero, termina el bucle y se envía de regreso a la llamada de función. El siguiente ejemplo ilustra cómo se puede aplicar el algoritmo genérico a un cálculo iterativo específico.

```
FUNCTION IterMeth(val, es, maxit)
iter = 1
sol = val
ea = 100
DO
  solold = sol
  sol = ...
  iter = iter + 1
  IF sol ≠ 0 ea=abs((sol - solold)/sol)*100
  IF ea ≤ es OR iter ≥ maxit EXIT
END DO
IterMeth = sol
END IterMeth
```

Figura 3.3
Pseudocódigo para un cálculo genérico iterativo.

EJEMPLO 3.3 Implementación en computadora de un cálculo iterativo

Planteamiento del problema Desarrolle un programa de computación basado en el pseudocódigo de la figura 3.3 para implementar el cálculo del ejemplo 3.2.

Solución Una función para implementar la expansión de la serie de Maclaurin para e^x se puede basar en el esquema general de la figura 3.3. Para hacer esto, primero se formula la expansión de la serie como una fórmula:

$$e^x \cong \sum_{i=0}^{n} \frac{x^n}{n!}$$

La figura 3.4 muestra las funciones para implementar esta serie, escritas en VBA y MATLAB. Se podrían desarrollar códigos similares en otros lenguajes como C++ o Fortran 95. Observe que, mientras MATLAB tiene una función `factorial` interna, es necesario calcular el factorial como parte de la implementación en VBA con un solo acumulador de productos `fac`.

Al correr los programas, generan un estimado para la función exponencial. Para la versión MATLAB, la respuesta se devuelve junto con el error aproximado y el número de iteraciones. Por ejemplo, e^1 se puede evaluar como

```
>> format long
>> [val, ea, iter] = IterMeth(1,1e-6,100)
val =
   2.718281826198493
ea =
     9.216155641522974e-007
iter =
     12
```

Se puede ver que, después de 12 iteraciones, se obtiene un resultado de 2.7182818 con un error aproximado de estimación = 9.2162×10^{-7}%. El resultado se puede verificar usando la función interna `exp` para calcular directamente el valor exacto y el error relativo porcentual verdadero.

```
>> trueval=exp(1)

trueval =
   2.718281828459046
```

```
>> et=abs((trueval-val)/trueval)*100
et =
    8.316108397236229e-008
```

Como en el caso del ejemplo 3.2, se obtuvo el resultado deseable de que el error verdadero es menor que el error aproximado.

a) VBA/Excel

```
Function IterMeth(x, es, maxit)
' initialization
iter = 1
sol = 1
ea = 100
fac = 1
' iterative calculation
Do
  solold = sol
  fac = fac * iter
  sol = sol + x ^ iter / fac
  iter = iter + 1
  If sol <> 0 Then
    ea = Abs((sol - solold) / sol) * 100
  End If
  If ea <= es Or iter >= maxit Then Exit Do
Loop
IterMeth = sol
End Function
```

b) MATLAB

```
function [v,ea,iter] = IterMeth(x,es,maxit)
% initialization
iter = 1;
sol = 1;
ea = 100;

% iterative calculation
while (1)
  solold = sol;

  sol = sol + x ^ iter / factorial(iter);
  iter = iter + 1;
  if sol~=0
    ea=abs((sol - solold)/sol)*100;
  end
  if ea<=es | iter>=maxit,break,end
end
v = sol;
end
```

Figura 3.4
a) VBA/Excel y *b*) Funciones MATLAB basadas en el pseudocódigo de la figura 3.3.

Con las definiciones anteriores como antecedente, se procede ahora a examinar los dos tipos de error relacionados directamente con los métodos numéricos: el error de redondeo y el error de truncamiento.

3.4 ERRORES DE REDONDEO

Como se mencionó antes, los errores de redondeo se originan debido a que la computadora emplea un número determinado de cifras significativas durante un cálculo. Los números como π, e o $\sqrt{7}$ no pueden expresarse con un número fijo de cifras significativas. Por lo tanto, no pueden ser representados exactamente por la computadora. Además, debido a que las computadoras usan una representación en base 2, no pueden representar exactamente algunos números en base 10. Esta discrepancia por la omisión de cifras significativas se llama *error de redondeo.*

3.4.1 Representación de números en la computadora

Numéricamente los errores de redondeo se relacionan de manera directa con la forma en que se guardan los números en la memoria de la computadora. La unidad fundamental mediante la cual se representa la información se llama *término.* Ésta es una entidad que consiste en una cadena de *dígitos binarios* o *bits* (*binary digits*). Por lo común, los números son guardados en una o más términos. Para entender cómo se realiza esto, se debe revisar primero algún material relacionado con los sistemas numéricos.

Sistemas numéricos Un *sistema numérico* es simplemente una convención para representar cantidades. Debido a que se tienen 10 dedos en las manos y 10 dedos en los pies, el sistema de numeración que nos es muy familiar es el *decimal* o de *base* 10. Una base es el número que se usa como referencia para construir un sistema. El sistema de base 10 utiliza 10 dígitos (0, 1, 2, 3, 4, 5, 6, 7, 8, 9) para representar números. Tales dígitos son satisfactorios por sí mismos para contar de 0 a 9.

Para grandes cantidades se usa la combinación de estos dígitos básicos; con la posición o *valor de posición* se especifica su magnitud. El dígito en el extremo derecho de un número entero representa un número del 0 al 9. El segundo dígito a partir de la derecha representa un múltiplo de 10. El tercer dígito a partir de la derecha representa un múltiplo de 100 y así sucesivamente. Por ejemplo, si se tiene el número 86 409 se tienen 8 grupos de 10 000, seis grupos de 1 000, cuatro grupos de 100 y cero grupos de 10, y nueve unidades, o bien

$$(8 \times 10^4) + (6 \times 10^3) + (4 \times 10^2) + (0 \times 10^1) + (9 \times 10^0) = 86\,409$$

La figura 3.5*a* ofrece una representación de cómo se formula un número en el sistema de base 10. Este tipo de representación se llama *notación posicional*.

Debido a que el sistema decimal resulta ser tan familiar, no es común darse cuenta de que existen otras alternativas. Por ejemplo, si el ser humano tuviera ocho dedos en las manos y ocho en los pies, se tendría, sin duda, una representación en un sistema *octal* o de *base* 8. En tal sentido nuestra amiga la computadora es como un animal que tiene dos dedos, limitado a dos estados: 0 o 1. Esto se relaciona con el hecho de que las unidades lógicas fundamentales de las computadoras digitales sean componentes electrónicos de apagado/encendido. Por lo tanto, los números en la computadora se representan con un sistema *binario* o de *base* 2. Del mismo modo que con el sistema decimal, las cantidades pueden representarse usando la notación posicional. Por ejemplo, el número binario 11 es equivalente a $(1 \times 2^1) + (1 \times 2^0) = 2 + 1 = 3$ en el sistema decimal. En la figura 3.5*b* se ilustra un ejemplo más complejo.

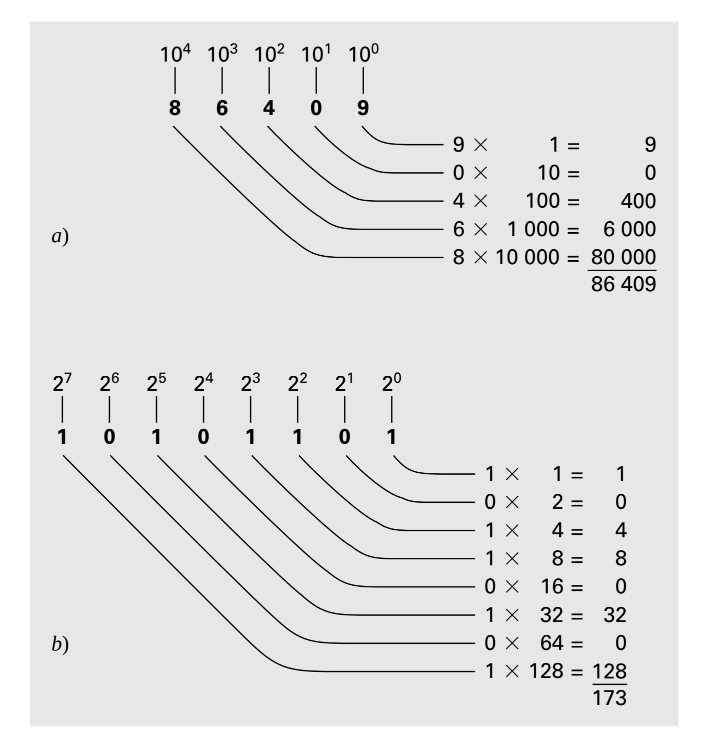


Figura 3.5
Cómo trabajan los sistemas *a*) decimal (base 10) y *b*) binario (base 2). En *b*) el número binario 10101101 es equivalente al número decimal 173.

Representación entera Ahora que se ha revisado cómo los números de base 10 se representan en forma binaria, es fácil concebir cómo los enteros se representan en la computadora. El método más sencillo se denomina *método de magnitud con signo* y emplea el primer bit de un término para indicar el signo: con un 0 para positivo y un 1 para el negativo. Los bits sobrantes se usan para guardar el número. Por ejemplo, el valor entero –173 puede guardarse en la memoria de una computadora de 16 bits como se muestra en la figura 3.6.

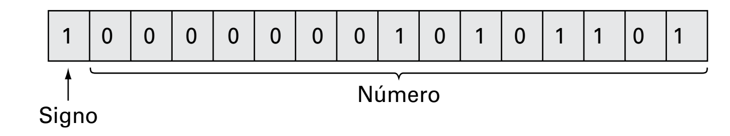


Figura 3.6
La representación de un entero decimal –173 en una computadora de 16 bits usando el método de magnitud con signo.

EJEMPLO 3.4 Rango de enteros

Planteamiento del problema Determine el rango de enteros de base 10 que pueda representarse en una computadora de 16 bits.

Solución De los 16 bits, se tiene el primer bit para el signo. Los 15 bits restantes pueden contener los números binarios de 0 a 111111111111111. El límite superior se convierte en un entero decimal, así

$$(1 \times 2^{14}) + (1 \times 2^{13}) + \cdots + (1 \times 2^1) + (1 \times 2^0)$$

que es igual a 32 767 (observe que esta expresión puede simplemente evaluarse como $2^{15} - 1$). Así, en una computadora de 16 bits un término puede guardar en memoria un entero decimal en el rango

de –32 767 a 32 767. Además, debido a que el cero está ya definido como 0000000000000000, sería redundante usar el número 1000000000000000 para definir “menos cero”. Por lo tanto, es usualmente empleado para representar un número negativo adicional: –32 768, y el rango va de –32 768 a 32 767.

Observe que el método de magnitud con signo descrito antes no se utiliza para representar enteros en computadoras convencionales. Se prefiere usar una técnica llamada *complemento de* 2 que incorpora en forma directa el signo dentro de la magnitud del número, en lugar de emplear un bit adicional para representar más o menos (véase Chapra y Canale, 1994). Sin embargo, el ejemplo 3.4 sigue sirviendo para ilustrar cómo todas las computadoras digitales están limitadas en cuanto a su capacidad para representar enteros. Esto es, los números por encima o por debajo de este rango no pueden representarse. Una limitación más importante se encuentra en el almacenaje y la manipulación de cantidades fraccionarias, como se describe a continuación.

Representación del punto flotante Las cantidades fraccionarias generalmente se representan en la computadora usando la forma de punto flotante. Con este método, el número se expresa como una parte fraccionaria, llamada *mantisa* o *significando*, y una parte entera, denominada *exponente* o *característica*, esto es,

$$m \cdot b^e$$

donde m = la mantisa, b = la base del sistema numérico que se va a utilizar y e = el exponente. Por ejemplo, el número 156.78 se representa como 0.15678×10^3 en un sistema de base 10 de punto flotante.

En 1a figura 3.7 se muestra una forma en que el número de punto flotante se guarda en una palabra. El primer bit se reserva para el signo; la siguiente serie de bits, para el exponente con signo; y los últimos bits, para la mantisa.

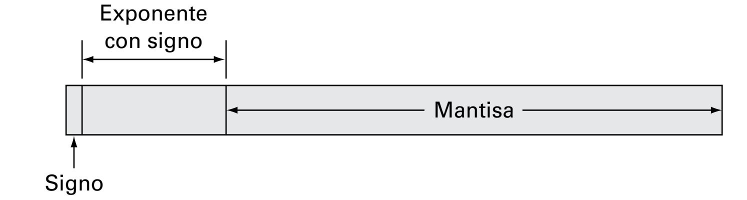


Figura 3.7
La forma en que un número de punto flotante se guarda en un término.

Observe que la mantisa es usualmente *normalizada* si tiene primero cero dígitos. Por ejemplo, suponga que la cantidad 1/34 = 0.029411765… se guarda en un sistema de base 10 con punto flotante, que únicamente permite guardar cuatro lugares decimales. Entonces, 1/34 se guardaría como

$$0.0294 \times 10^0$$

Sin embargo, al hacerlo así, la inclusión del cero “inútil” a la derecha del punto decimal nos obliga a eliminar el dígito 1 del quinto lugar decimal. El número puede normalizarse para eliminar el cero multiplicando la mantisa por 10 y disminuyendo el exponente en 1, para quedar

$$0.2941 \times 10^{-1}$$

Así, se conserva una cifra significativa adicional al guardar el número.

La consecuencia de la normalización es que el valor absoluto de m queda limitado. Esto es,

$$\frac{1}{b} \leq m < 1 \tag{3.8}$$

donde b = la base. Por ejemplo, para un sistema de base 10, m estaría entre 0.1 y 1; y para un sistema de base 2, entre 0.5 y 1.

La representación de punto flotante permite que tanto fracciones como números muy grandes se expresen en la computadora. Sin embargo, hay algunas desventajas. Por ejemplo, los números de punto flotante requieren más espacio y más tiempo de procesado que los números enteros. Más im-

portante aún es que su uso introduce una fuente de error debido a que la mantisa conserva sólo un número finito de cifras significativas. Por lo tanto, se introduce un error de redondeo.

EJEMPLO 3.5 Conjunto hipotético de números con punto flotante

Planteamiento del problema Determine un conjunto hipotético de números con punto flotante para una máquina que guarda información usando términos de 7 bits. Emplee el primer bit para el signo del número, los siguientes tres para el signo y la magnitud del exponente, y los últimos tres para la magnitud de la mantisa (véase figura 3.8).

Solución El número positivo más pequeño posible se representa en la figura 3.8. El 0 inicial señala que la cantidad es positiva. El 1 en el segundo lugar indica que el exponente tiene signo negativo. Los 1, en el tercero y cuarto lugar dan un valor máximo al exponente de

$$1 \times 2^1 + 1 \times 2^0 = 3$$

Por lo tanto, el exponente será –3. Por último, la mantisa está especificada por el 100 en los últimos tres lugares, lo cual nos da

$$1 \times 2^{-1} + 0 \times 2^{-2} + 0 \times 2^{-3} = 0.5$$

Figura 3.8
El número positivo de punto flotante más pequeño posible del ejemplo 3.5.

Aunque es posible tomar una mantisa más pequeña (por ejemplo, 000, 001, 010, 011), se emplea el valor de 100 debido al límite impuesto por la normalización [ecuación (3.8)]. Así, el número positivo más pequeño posible en este sistema es $+0.5 \times 2^{-3}$, el cual es igual a 0.0625 en el sistema de base 10. Los siguientes números más grandes se desarrollan incrementando la mantisa como sigue:

$$0111101 = (1 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3}) \times 2^{-3} = (0.078125)_{10}$$
$$0111110 = (1 \times 2^{-1} + 1 \times 2^{-2} + 0 \times 2^{-3}) \times 2^{-3} = (0.093750)_{10}$$
$$0111111 = (1 \times 2^{-1} + 1 \times 2^{-2} + 1 \times 2^{-3}) \times 2^{-3} = (0.109375)_{10}$$

Observe que las equivalencias de base 10 se esparcen de manera uniforme en un intervalo de 0.015625. En este punto, para continuar el incremento se debe disminuir el exponente a 10, lo cual da un valor de

$$1 \times 2^1 + 0 \times 2^0 = 2$$

La mantisa disminuye hasta su valor más pequeño: 100. Por lo tanto, el siguiente número es

$$0110100 = (1 \times 2^{-1} + 0 \times 2^{-2} + 0 \times 2^{-3}) \times 2^{-2} = (0.125000)_{10}$$

Esto todavía representa una brecha o espacio de 0.l25000 – 0.109375 = 0.015625. Sin embargo, cuando los números grandes se generan incrementando la mantisa, la brecha es de 0.03125,

$$0110101 = (1 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3}) \times 2^{-2} = (0.156250)_{10}$$
$$0110110 = (1 \times 2^{-1} + 1 \times 2^{-2} + 0 \times 2^{-3}) \times 2^{-2} = (0.187500)_{10}$$
$$0110111 = (1 \times 2^{-1} + 1 \times 2^{-2} + 1 \times 2^{-3}) \times 2^{-2} = (0.218750)_{10}$$

Este patrón se repite conforme se formula una cantidad mayor hasta que se alcanza un número máximo:

$$0011111 = (1 \times 2^{-1} + 1 \times 2^{-2} + 1 \times 2^{-3}) \times 2^{3} = (7)_{10}$$

El conjunto del número final se muestra en la figura 3.9.

En la figura 3.9 se presentan diversos aspectos de la representación de punto flotante, que son importantes respecto de los errores de redondeo en las computadoras.

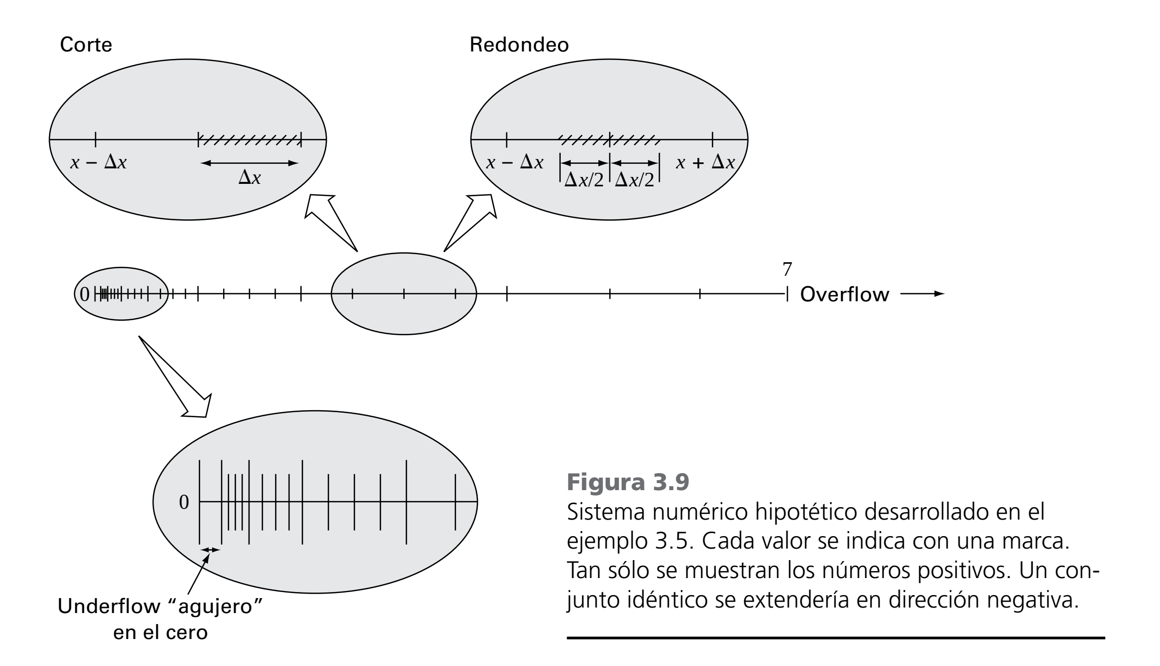


Figura 3.9
Sistema numérico hipotético desarrollado en el ejemplo 3.5. Cada valor se indica con una marca. Tan sólo se muestran los números positivos. Un conjunto idéntico se extendería en dirección negativa.

1. *El rango de cantidades que pueden representarse es limitado.* Como en el caso de los enteros, hay números grandes positivos y negativos que no pueden representarse. Intentar emplear números fuera del rango aceptable dará como resultado el llamado *error de desbordamiento* (*overflow*). Sin embargo, además de las grandes cantidades, la representación de punto flotante tiene la limitación adicional de que números muy pequeños no pueden representarse. Esto se ilustra por el "agujero" *underflow* entre el cero y el primer número positivo en la figura 3.9. Se debe observar que este agujero aumenta por las limitaciones de normalización de la ecuación (3.8).
2. *Existe sólo un número finito de cantidades que puede representarse dentro de un rango.* Así, el grado de precisión es limitado. Es evidente que los números irracionales no pueden representarse de manera exacta. Además, los números racionales que no concuerdan exactamente con uno de los valores en el conjunto tampoco pueden ser representados en forma precisa. A los errores ocasionados por la aproximación en ambos casos se les conoce como errores de *cuantificación.* La aproximación real se realiza por dos caminos: cortando o redondeando. Por ejemplo, suponga que el valor de π = 3.14159265358… se va a guardar en un sistema de numeración de base 10 con 7 cifras significativas. Un método de aproximación podría ser simplemente omitir, o "cortar", el octavo y demás términos, como en π = 3.141592, con la introducción de un error asociado de [ecuación (3.2)]

   $$E_t = 0.00000065\ldots$$

   Esta técnica de mantener sólo términos significativos fue originalmente conocida como "truncamiento" en la jerga computacional. Preferimos llamarla *corte* para distinguirla de los errores de truncamiento que se analizarán en el capítulo 4. Observe que en el sistema numérico de base 2 de la figura 3.9, corte significa que cualquier cantidad que esté dentro de un intervalo de longitud Δx se guardará en memoria como una cantidad en el extremo inferior del intervalo. Así, el error máximo por corte es Δx. Además, se presenta un sesgo porque todos los errores son positivos. La deficiencia del corte se atribuye al hecho de que los términos superiores de la representación decimal completa no tienen impacto en la versión cortada. Así, en el ejemplo de π, el primer dígito descartado es 6. El último dígito retenido debería redondearse a 3.141593. Tal *redondeo* reduce el error a

   $$E_t = -0.00000035\ldots$$

   En consecuencia, el redondeo produce un error absoluto menor que el de corte. Observe que, en el sistema numérico de base 2 de la figura 3.9, redondear significa que cualquier cantidad que esté en un intervalo de longitud Δx se representará como el número más cercano permitido.

Entonces, el error máximo de redondeo es $\Delta x/2$. Además, no se presenta sesgo porque ciertos errores son positivos y otros son negativos. Algunas computadoras emplean redondeo. Sin embargo, esto aumenta el trabajo computacional y, en consecuencia, muchas máquinas simplemente usan el corte. Dicho enfoque se justifica con la suposición de que el número de cifras significativas es suficientemente grande para que los errores de redondeo resultantes sean despreciables.

3. *El intervalo entre los números, Δx, aumenta conforme los números crecen en magnitud*. Ésta es la característica, por supuesto, que permite que la representación de punto flotante conserve los dígitos significativos. Sin embargo, también quiere decir que los errores de cuantificación sean proporcionales a la magnitud del número que será representado. Para normalizar los números de punto flotante, esta proporcionalidad se expresa, para los casos en que se emplea el corte, como

$$\frac{|\Delta x|}{|x|} \leq \mathscr{E} \tag{3.9}$$

y, para los casos donde se utiliza el redondeo, como

$$\frac{|\Delta x|}{|x|} \leq \frac{\mathscr{E}}{2} \tag{3.10}$$

donde a $\mathscr{E}$ se le denomina *épsilon de la máquina*, el cual se calcula como

$$\mathscr{E} = b^{1-t} \tag{3.11}$$

donde b es el número base y t es el número de dígitos significativos en la mantisa. Observe que las desigualdades en las ecuaciones (3.9) y (3.10) quieren decir que éstos son los límites de los errores. Es decir, especifican los casos extremos.

EJEMPLO 3.6 Épsilon de la máquina

Planteamiento del problema Determine el épsilon de la máquina y verifique su efectividad para caracterizar los errores del sistema numérico del ejemplo 3.5. Suponga que se usa al corte.

Solución El sistema de punto flotante hipotético del ejemplo 3.5 empleaba valores de base $b = 2$, y número de bits de la mantisa $t = 3$. Por lo tanto, el épsilon de la máquina debe ser [ecuación (3.11)]

$$\mathscr{E} = 2^{1-3} = 0.25$$

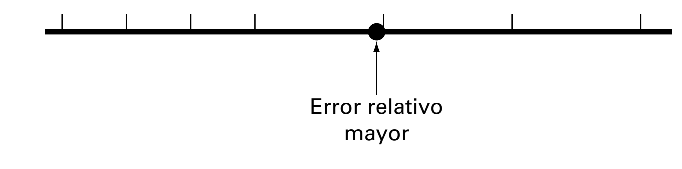


Figura 3.10
El error de cuantificación más grande ocurrirá para aquellos valores que caigan justo debajo del límite superior del primero de una serie de intervalos equiespaciados.

En consecuencia, el error de cuantificación relativo estará limitado por 0.25, para el corte. El error relativo más grande debería ocurrir para aquellas cantidades que caen justo debajo del límite superior del primer intervalo entre números equidistantes sucesivos (véase figura 3.10). Aquellos números que caen en los intervalos sucesivos siguientes tendrán el mismo valor de Δx pero un mayor valor de x y, por lo tanto, tendrán un error relativo bajo. Un ejemplo de un error máximo sería un valor que cae justo por debajo de límite superior del intervalo entre $(0.125000)_{10}$ y $(0.156250)_{10}$. Para este caso, el error sería menor a

$$\frac{0.03125}{0.125000} = 0.25$$

Entonces, el error es como se predijo mediante la ecuación (3.9).

El hecho de que los errores de cuantificación dependan de la magnitud tiene varias aplicaciones prácticas en los métodos numéricos. Muchas de ellas están relacionadas con la comúnmente empleada operación de probar si dos números son iguales. Ello ocurre cuando se prueba la convergencia de cantidades, así como en los mecanismos para detener procesos iterativos (véase el ejemplo 3.2). En estos casos deberá ser claro que más que probar si las dos cantidades son iguales, es recomendable probar si su diferencia es menor que una pequeña tolerancia aceptable. Además, deberá ser evidente que más que la diferencia absoluta, deberá compararse la diferencia normalizada, en especial cuando se trabaja con números de gran magnitud. El épsilon de la máquina, además, se emplea al formular criterios de paro o de convergencia. Esto asegura que los programas sean portátiles, es decir, que no sean dependientes de la computadora sobre la cual se hayan implementado. En la figura 3.11 se presenta un pseudocódigo que automáticamente determina el épsilon de la máquina en una computadora binaria.

```
epsilon = 1
DO
  IF (epsilon+1 ≤ 1) EXIT
  epsilon = epsilon/2
END DO
epsilon = 2 × epsilon
```

Figura 3.11
Pseudocódigo para determinar el épsilon de la máquina en una computadora binaria.

Precisión extendida Aquí se debe observar que, aunque los errores de redondeo llegan a ser importantes en contextos como pruebas de convergencia, el número de dígitos significativos que tiene la mayoría de las computadoras permite que muchos cálculos de ingeniería se realicen con una precisión más que aceptable. Por ejemplo, el sistema numérico hipotético de la figura 3.9 es una enorme exageración que se usó con propósitos ilustrativos. En las computadoras comerciales se utilizan conjuntos mucho más grandes y por consiguiente se permite que los números queden expresados con una precisión adecuada. Por ejemplo, las computadoras que usan el formato IEEE permiten 24 bits para ser usados por la mantisa, lo cual se traduce en cerca de siete cifras significativas de precisión[1] en dígitos de base 10 con un rango aproximado de 10^{-38} a 10^{39}.

Se debe reconocer que aún hay casos donde el error de redondeo resulta crítico. Por tal razón muchas computadoras permiten la especificación de precisión extendida. La más común de estas especificaciones es la doble precisión, en la cual se duplica el número de términos utilizado para guardar números de punto flotante. Esto proporciona de 15 a 16 dígitos decimales de precisión y un rango aproximado de 10^{-308} a 10^{308}.

En muchos casos el uso de cantidades de doble precisión llega a reducir, en gran medida, el efecto del error de redondeo. Sin embargo, el precio que se paga por tales medidas remediales consiste en mayores requerimientos de memoria y de tiempo de ejecución. La diferencia en el tiempo de ejecución de un cálculo pequeño podría parecer insignificante. No obstante, conforme los programas van siendo cada vez más grandes y complicados, el tiempo de ejecución agregado se vuelve más considerable y repercute de manera negativa para resolver el problema en forma efectiva. Por lo tanto, la precisión extendida no debería utilizarse en forma generalizada. Por el contrario, deberá ser empleada en forma selectiva, donde se obtenga un máximo beneficio al menor costo en términos de tiempo de ejecución. En las siguientes secciones veremos más de cerca cómo los errores de redondeo afectan los cálculos y ello nos servirá para comprender los fundamentos que nos guíen en el uso de la capacidad de la doble precisión.

Antes de proseguir, debemos observar que algunos paquetes de software de uso común (por ejemplo, Excel o Mathcad) normalmente utilizan doble precisión para representar las cantidades numéricas. Así, quienes desarrollaron estos paquetes decidieron reducir los errores de redondeo sacrificando velocidad para usar una precisión extendida. Otros, como el MATLAB, permiten usar la precisión extendida, si se desea.

3.4.2 Manipulación aritmética de números en la computadora

Junto con las limitaciones del sistema numérico de una computadora, las manipulaciones aritméticas que se usan con tales números también pueden dar como resultado errores de redondeo. En la siguiente sección se ilustrará primero cómo afectan las operaciones aritméticas comunes a los errores

---

[1] Observe que, de hecho, únicamente 23 bits se emplean en la memoria para la mantisa. Sin embargo, debido a la normalización, el primer bit de la mantisa es siempre 1 y, por lo tanto, no se guarda. Así, el primer bit junto con los 23 bits de memoria dan 24 bits en total para la precisión de la mantisa.

de redondeo. De este modo, investigaremos varias manipulaciones que son especialmente propensas a errores de redondeo.

Operaciones aritméticas comunes A causa de que estamos familiarizados con los números de base 10, los emplearemos para ilustrar el efecto del error de redondeo en las operaciones básicas: suma, resta, multiplicación y división. Otras bases de números pueden tener un comportamiento similar. Para simplificar el análisis, emplearemos una computadora decimal hipotética con una mantisa de 4 dígitos y un exponente de 1 dígito. Además, se usará el corte. El redondeo puede implicar errores similares, aunque menos dramáticos.

Cuando se suman dos números de punto flotante, el número de la mantisa con el exponente menor se modifica de tal forma que los exponentes sean los mismos. Esto tiene el efecto de alinear los puntos decimales. Por ejemplo, suponga que se quiere sumar $0.1557 \cdot 10^1 + 0.4381 \cdot 10^{-1}$. El decimal de la mantisa del segundo número se recorre a la izquierda un número de lugares igual a la diferencia de los exponentes [1 – (–1) = 2], así,

$$0.4381 \cdot 10^{-1} \rightarrow 0.004381 \cdot 10^1$$

Ahora se suman los números,

$$\begin{array}{l} 0.1557 \quad\;\; \cdot 10^1 \\ \underline{0.004381 \cdot 10^1} \\ 0.160081 \cdot 10^1 \end{array}$$

y el resultado es cortado a $0.1600 \cdot 10^1$. Note cómo los últimos dos dígitos del segundo número que se recorrieron a la derecha fueron eliminados de los cálculos.

La resta se realiza en forma idéntica a la suma, con la excepción del signo del sustraendo, que es negativo. Por ejemplo, suponga que hacemos la resta 36.41 menos 26.86. Esto es,

$$\begin{array}{rl} & 0.3641 \cdot 10^2 \\ - & \underline{0.2686 \cdot 10^2} \\ & 0.0955 \cdot 10^2 \end{array}$$

Aquí el resultado no está normalizado y se debe recorrer el decimal un lugar a la derecha para obtener $0.9550 \cdot 10^1 = 9.550$. Observe que el cero sumado al final de la mantisa no es relevante, tan sólo llena el espacio vacío creado al recorrer los números. Es posible obtener resultados más dramáticos todavía, cuando las cantidades estén muy cercanas, como por ejemplo,

$$\begin{array}{rl} & 0.7642 \cdot 10^3 \\ - & \underline{0.7641 \cdot 10^3} \\ & 0.0001 \cdot 10^3 \end{array}$$

que podría convertirse en $0.1000 \cdot 10^0 = 0.1000$. Así, en este caso, se agregan tres ceros no significativos, lo cual introduce un error sustancial de cálculo debido a que las manipulaciones siguientes actúan como si los ceros fueran significativos. Como se verá más adelante en otra sección, la pérdida significativa durante la resta de números casi iguales es una de las principales fuentes de errores de redondeo en los métodos numéricos.

La multiplicación y la división resultan un poco más sencillos que la suma y la resta. Los exponentes se suman y la mantisa se multiplica. Debido a que la multiplicación de dos mantisas de n dígitos da como resultado $2n$ dígitos, muchas computadoras ofrecen resultados intermedios en un registro de doble longitud. Por ejemplo,

$$0.1363 \cdot 10^3 \times 0.6423 \cdot 10^{-1} = 0.08754549 \cdot 10^2$$

Si, como en este caso, se introduce un cero, el resultado es normalizado,

$$0.08754549 \cdot 10^2 \rightarrow 0.8754549 \cdot 10^1$$

y cortando resulta

$$0.8754 \cdot 10^1$$

La división se realiza en forma similar, aunque las mantisas se dividen y los exponentes se restan. Entonces el resultado es normalizado y cortado.

Cálculos grandes Ciertos métodos requieren un número extremadamente grande de manipulaciones aritméticas para llegar a los resultados finales. Además, dichos cálculos a menudo son interdependientes; es decir, los cálculos son dependientes de los resultados previos. En consecuencia, aunque el error de redondeo individual sea pequeño, el efecto acumulativo durante el proceso de muchos cálculos puede ser relevante.

EJEMPLO 3.7 Un número grande de cálculos interdependientes

Planteamiento del problema Investigue el efecto del error de redondeo en un gran número de cálculos interdependientes. Desarrolle un programa que sume un número 100 000 veces. Sume el número 1 con precisión simple, y 0.00001 con precisiones simple y doble.

Solución En la figura 3.12 se muestra un programa en Fortran 90 que realiza la suma. Mientras que la suma con precisión simple de 1 dará el resultado esperado, la precisión simple en la suma de 0.00001 tiene una gran discrepancia. Este error se reduce de manera importante cuando 0.00001 se suma con precisión doble.

Los errores de cuantificación son la fuente de las discrepancias. Debido a que el entero 1 puede ser representado en forma exacta en la computadora, puede sumarse exactamente. En contraste, 0.00001 no puede representarse con exactitud y se cuantifica con un valor que es ligeramente diferente de su valor verdadero. Aunque esta ligera discrepancia resultará insignificante para un cálculo pequeño, se acumula después de la repetición de sumas. Tal problema ocurre también con la precisión doble, pero se reduce en forma relevante porque el error de cuantificación es mucho más pequeño.

```
PROGRAM fig0312
IMPLICIT none
INTEGER::i
REAL::sum1, sum2, x1, x2
DOUBLE PRECISION::sum3, x3
sum1=0.
sum2=0.
sum3=0.
x1=1.
x2=1.e-5
x3=1.d-5
DO i=1,100000
  sum1=sum1+x1
  sum2=sum2+x2
  sum3=sum3+x3
END DO
PRINT *, sum1
PRINT *, sum2
PRINT *, sum3
END
output:
100000.000000
         1.000990
  9.999999999980838E-001
```

Figura 3.12
Programa en Fortran 90 para sumar un número 10^5 veces. Aquí se suma el número 1 con precisión simple y el número 10^{-5} con precisiones simple y doble.

Observe que el tipo de error ilustrado en el ejemplo anterior es algo atípico porque todos los errores en las operaciones que se repiten tienen el mismo signo. En muchos casos, los errores en grandes cálculos alternan el signo de manera aleatoria y, entonces, con frecuencia se cancelan. Sin embargo, hay también algunos casos donde tales errores no se cancelan pero, en efecto, llevan a resultados finales dudosos. En las siguientes secciones se mostrará cómo puede ocurrir esto.

Suma de un número grande y uno pequeño Suponga que se desea sumar un número pequeño, 0.0010, con un número grande, 4 000, utilizando una computadora hipotética con una mantisa de 4 dígitos y un exponente de 1 dígito. Modificamos el número pequeño para que su exponente sea igual al del grande,

$$\begin{array}{l} 0.4000 \quad\;\;\;\, \cdot 10^4 \\ \underline{0.0000001 \cdot 10^4} \\ 0.4000001 \cdot 10^4 \end{array}$$

el cual se corta a $0.4000 \cdot 10^4$. Así, ¡resultó lo mismo que si no hubiéramos realizado la suma!

Este tipo de error puede ocurrir cuando se calculan series infinitas. Por ejemplo, si el término inicial de una serie es relativamente grande en comparación con los demás términos, después de que se han sumado unos pocos términos, estamos en la situación de sumar una cantidad pequeña a una cantidad grande.

Una manera de reducir este tipo de errores consiste en sumar la serie en sentido inverso: esto es, en orden ascendente en lugar de descendente. De esta manera, cada nuevo término será comparable en magnitud con el de la suma acumulada (véase el problema 3.5).

Cancelación por resta Se refiere al redondeo inducido cuando se restan dos números de punto flotante casi iguales.

Un caso común donde esto ocurre es en la determinación de las raíces de una ecuación cuadrática o parábola utilizando la fórmula cuadrática,

$$\begin{matrix} x_1 \\ x_2 \end{matrix} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{3.12}$$

En los casos donde $b^2 \gg 4ac$, la diferencia en el numerador puede ser muy pequeña. En tales casos, la precisión doble llega a reducir el problema. Además, una formulación alternativa puede usarse para minimizar la cancelación por resta.

$$\begin{matrix} x_1 \\ x_2 \end{matrix} = \frac{-2c}{b \pm \sqrt{b^2 - 4ac}} \tag{3.13}$$

Una ilustración del problema y del uso de esta fórmula alternativa se ofrecen en el siguiente ejemplo.

```
Option Explicit

Sub fig0313()
Dim a As Single, b As Single
Dim c As Single, d As Single
Dim x1 As Single, x2 As Single
Dim x1r As Single
Dim aa As Double, bb As Double
Dim cc As Double, dd As Double
Dim x11 As Double, x22 As Double

'Single precision:
a = 1: b = 3000.001: c = 3
d = Sqr(b * b - 4 * a * c)
x1 = (-b + d) / (2 * a)
x2 = (-b - d) / (2 * a)

'Double precision:
aa = 1: bb = 3000.001: cc = 3
dd = Sqr(bb * bb - 4 * aa * cc)
x11 = (-bb + dd) / (2 * aa)
x22 = (-bb - dd) / (2 * aa)

'Modified formula for first root
'single precision:
x1r = -2 * c / (b + d)

'Display results
Sheets("sheet1").Select
Range("b2").Select
ActiveCell.Value = x1
ActiveCell.Offset(1, 0).Select
ActiveCell.Value = x2
ActiveCell.Offset(2, 0).Select
ActiveCell.Value = x11
ActiveCell.Offset(1, 0).Select
ActiveCell.Value = x22
ActiveCell.Offset(2, 0).Select
ActiveCell.Value = x1r
End Sub
```

OUTPUT:

	A	B	C
1	Single-precision results:		
2	x1	-0.000976563	
3	x2	-3000.00000000	
4	Double-precision results:		
5	x1	-0.00100000	
6	x2	-3000.00000000	
7	Modified formula for first root (single precision):		
8	x1	-0.00100000	

Figura 3.13
Programa Excel/VBA para determinar las raíces de una cuadrática.

EJEMPLO 3.8 Cancelación por resta

Planteamiento del problema Calcule el valor de las raíces de una ecuación cuadrática con $a = 1$, $b = 3000.001$ y $c = 3$. Compare el valor calculado con las raíces verdaderas $x_1 = -0.001$ y $x_2 = -3000$.

Solución En la figura 3.13 se muestra un programa en Excel/VBA que calcula las raíces x_1 y x_2 usando la fórmula cuadrática (ecuación 3.12). Observe que se dan las versiones tanto de la precisión simple como la precisión doble. Mientras que los resultados para x_2 son adecuados, el error relativo porcentual para x_1 es pobre para la precisión simple, $\varepsilon_t = 2.4\%$. Este valor quizá resulte inadecuado para muchos problemas de aplicaciones en ingeniería. ¡Este resultado es en particular sorpresivo porque se emplea una fórmula analítica para obtener la solución!

La pérdida de significancia ocurre en la línea del programa donde dos números grandes se restan. No ocurren problemas semejantes cuando los mismos números se suman.

Considerando lo anterior podemos obtener la conclusión general de que la fórmula cuadrática será susceptible de cancelación por resta cada vez que $b^2 \gg 4ac$. Una manera de evitar este problema consiste en usar precisión doble. Otra es reacomodar la fórmula cuadrática en la forma de la ecuación (3.13). Ya que en la salida del programa, ambas opciones dan un error mucho menor porque se minimiza o evita la cancelación por resta.

Considere que, como en el ejemplo anterior, hay veces en las que la cancelación por resta se evita empleando una transformación. No obstante, el único remedio general es usar la precisión extendida.

Dispersión La dispersión ocurre generalmente cuando los términos individuales en la sumatoria son más grandes que la sumatoria misma. Como en el siguiente ejemplo, casos como éstos ocurren en las series con signos alternados.

EJEMPLO 3.9 Evaluación de e^x usando series infinitas

Planteamiento del problema La función exponencial $y = e^x$ está dada por la serie infinita

$$y = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \cdots$$

Evalúe esta función para $x = 10$ y $x = -10$; esté atento al problema del error de redondeo.

Solución En la figura 3.14*a* se muestra un programa en Excel/VBA que utiliza una serie infinita para evaluar e^x. La variable *i* es el número de términos en la serie, *term* es el valor de término actual que se le agrega a la serie, y *sum* es el valor acumulado de la serie. La variable *test* es el valor acumulado precedente de la serie antes de la suma de *term*. La serie se termina cuando la computadora no puede detectar la diferencia entre *test* y *sum*.

La figura 3.14*b* muestra los resultados de la ejecución del programa para $x = 10$. Observe que este caso es completamente satisfactorio. El resultado final se alcanza en 31 términos con la serie idéntica para el valor de la función en la biblioteca con siete cifras significativas.

En la figura 3.14*c* se muestran los resultados para $x = -10$. Sin embargo, en este caso, los resultados de la serie calculada no coinciden ni en el signo con respecto al resultado verdadero. De hecho, los resultados negativos abren una gama de preguntas serias porque e^x nunca puede ser menor que cero. El problema es causado por el error de redondeo. Observe que muchos de los términos que conforman la suma son mucho más grandes que el resultado final de la suma. Además, a diferencia del caso anterior, los términos individuales varían de signo. Así, en efecto, estamos sumando y restando números grandes (cada uno con algún error pequeño) y dando gran significancia a las diferencias; esto es, cancelación por resta. Entonces, puede verse que el culpable en este ejemplo de dispersión es, en efecto, la cancelación por resta. En tales casos es apropiado buscar alguna otra estrategia de cálculo. Por ejemplo, uno podría tratar de calcular $y = e^{10}$ como $y = (e^{-1})^{10}$. En lugar de una reformulación, ya que el único recurso general es la precisión extendida.

a) Programa

```
Option Explicit

Sub fig0314()
Dim term As Single, test As Single
Dim sum As Single, x As Single
Dim i As Integer
i = 0: term = 1#: sum = 1#: test = 0#
Sheets("sheet1").Select
Range("b1").Select
x = ActiveCell.Value
Range("a3:c1003").ClearContents
Range("a3").Select
Do
  If sum = test Then Exit Do
  ActiveCell.Value = i
  ActiveCell.Offset(0, 1).Select
  ActiveCell.Value = term
  ActiveCell.Offset(0, 1).Select
  ActiveCell.Value = sum
  ActiveCell.Offset(1, -2).Select
  i = i + 1
  test = sum
  term = x ^ i / _
    Application.WorksheetFunction.Fact(i)
  sum = sum + term
Loop
ActiveCell.Offset(0, 1).Select
ActiveCell.Value = "Exact value = "
ActiveCell.Offset(0, 1).Select
ActiveCell.Value = Exp(x)
End Sub
```

b) Evaluación de e^{10}

	A	B	C
1	x	10	
2	i	term	sum
3	0	1.000000	1.000000
4	1	10.000000	11.000000
5	2	50.000000	61.000000
6	3	166.666672	227.666672
7	4	416.666656	644.333313
8	5	833.333313	1477.666626
30	27	9.183690E-02	22026.416016
31	28	3.279889E-02	22026.449219
32	29	1.130996E-02	22026.460938
33	30	3.769988E-03	22026.464844
34	31	1.216125E-03	22026.466797
35		Exact value =	22026.465795

c) Evaluación de e^{-10}

	A	B	C
1	x	-10	
2	i	term	sum
3	0	1.000000	1.000000
4	1	-10.000000	-9.000000
5	2	50.000000	41.000000
6	3	-166.666672	-125.666672
7	4	416.666656	291.000000
8	5	-833.333313	-542.333313
44	41	-2.989311E-09	1.103359E-04
45	42	7.117407E-10	1.103366E-04
46	43	-1.655211E-10	1.103365E-04
47	44	3.761843E-11	1.103365E-04
48	45	-8.359651E-12	1.103365E-04
49		Exact value =	4.539993E-05

Figura 3.14
a) Un programa Excel/VBA para evaluar e^x usando una serie infinita. *b*) Evaluación de e^x. *c*) Evaluación de e^{-x}.

Productos internos De las secciones anteriores debe quedar claro que, algunas series infinitas son particularmente propensas a errores por redondeo. Por fortuna, el cálculo de series no es una de las operaciones más comunes en métodos numéricos. Una manipulación más frecuente es el cálculo de productos internos, esto es,

$$\sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

Esta operación es muy común, en particular en la solución de ecuaciones simultáneas lineales algebraicas. Tales sumatorias son propensas a errores por redondeo. En consecuencia, a menudo es deseable calcular tales sumas con precisión extendida.

Aunque en las secciones siguientes se ofrecerán reglas prácticas para reducir el error de redondeo, no son un medio directo mejor que el método de prueba y error para determinar realmente el efecto de tales errores en los cálculos. En el próximo capítulo se presentará la serie de Taylor, la cual proporcionará un enfoque matemático para estimar esos efectos.

PROBLEMAS

3.1 Convierta los números que se presentan base 2 a base 10: *a*) 101101; *b*) 110.011 y *c*) 0.01101.

3.2 Convierta los siguientes números base 8 a base 10: 71 263 y 3.147.

3.3 Realice su propio programa con base en la figura 3.11 y úselo para determinar el épsilon de máquina de su computadora.

3.4 En forma similar a la de la figura 3.11, escriba un programa corto para determinar el número más pequeño, $x_{mín}$, que utiliza la computadora que empleará con este libro. Observe que su computadora será incapaz de diferenciar entre cero y una cantidad más pequeña que dicho número.

3.5 La serie infinita

$$f(n) = \sum_{i=1}^{n} \frac{1}{i^4}$$

converge a un valor de $f(n) = \pi^4/90$ conforme n se tiende a infinito. Escriba un programa de precisión sencilla para calcular $f(n)$ para n = 10 000 por medio de calcular la suma desde i = 1 hasta 10 000. Después repita el cálculo pero en sentido inverso, es decir, desde i = 10 000 a 1, con incrementos de –1. En cada caso, calcule el error relativo porcentual verdadero. Explique los resultados.

3.6 Evalúe e^{-5} con el uso de dos métodos

$$e^{-x} = 1 - x + \frac{x^2}{2} - \frac{x^3}{3!} + \cdots$$

y

$$e^{-x} = \frac{1}{e^x} = \frac{1}{1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \cdots}$$

y compárelo con el valor verdadero de 6.737947×10^{-3}. Utilice 20 términos para evaluar cada serie y calcule los errores relativos aproximado y verdadero como términos que se agregaran.

3.7 La derivada de $f(x) = 1/(1 - 3x^2)$ está dada por

$$\frac{6x}{(1-3x^2)^2}$$

¿Esperaría dificultades para evaluar esta función para x = 0.577? Inténtelo con aritmética de 3 y 4 dígitos con corte.

3.8 *a*) Evalúe el polinomio

$$y = x^3 - 5x^2 + 6x + 0.55$$

en x = 1.37. Utilice aritmética de 3 dígitos con corte. Evalúe el error relativo porcentual.

b) Repita el inciso *a*) pero exprese a y como

$$y = ((x - 5)x + 6)x + 0.55$$

Evalúe el error y compárelo con el inciso *a*).

3.9 Calcule la memoria de acceso aleatorio (RAM) en megabytes, que es necesaria para almacenar un arreglo multidimensional de 20 × 40 × 120. Este arreglo es de doble precisión, y cada valor requiere una palabra de 64 bits. Recuerde que una palabra de 64 bits = 8 bytes, y un kilobyte = 2^{10} bytes. Suponga que el índice comienza en 1.

3.10 Determine el número de términos necesarios para aproximar cos x a 8 cifras significativas con el uso de la serie de Maclaurin.

$$\cos x = 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \cdots$$

Calcule la aproximación en donde el valor de $x = 0.3\pi$. Escriba un programa para determinar el resultado.

3.11 Utilice aritmética de 5 dígitos con corte para determinar las raíces de la ecuación siguiente, por medio de las ecuaciones (3.12) y (3.13).

$$x^2 - 5\,000.002x + 10$$

Calcule los errores relativos porcentuales de sus resultados.

3.12 ¿Cómo puede emplearse el épsilon de la máquina para formular un criterio de detención ε_s para sus programas? Dé un ejemplo.

3.13 El método “dividir y promediar”, un antiguo método para aproximar la raíz cuadrada de cualquier número positivo a se puede formular como

$$x = \frac{x + a/x}{2}$$

Escriba una función bien estructurada para implementar este algoritmo basado en el algoritmo descrito en la figura 3.3.

CAPÍTULO 4

Errores de truncamiento y la serie de Taylor

Los *errores de truncamiento* son aquellos que resultan al usar una aproximación en lugar de un procedimiento matemático exacto. Por ejemplo, en el capítulo 1 aproximamos la derivada de la velocidad de caída de un paracaidista mediante una ecuación en diferencias finitas divididas de la forma [ecuación (1.11)]

$$\frac{dv}{dt} \cong \frac{\Delta v}{\Delta t} = \frac{v(t_{i+1}) - v(t_i)}{t_{i+1} - t_i} \tag{4.1}$$

Se presentó un error de truncamiento en la solución numérica, ya que la ecuación en diferencia sólo se aproxima al valor exacto de la derivada (véase figura 1.4). Para obtener un conocimiento sobre las características de estos errores, debe considerar una formulación matemática que se utiliza ampliamente en los métodos numéricos para expresar funciones de manera aproximada: la serie de Taylor.

4.1 LA SERIE DE TAYLOR

El teorema de Taylor (véase cuadro 4.1) y su fórmula, la serie de Taylor, es de gran valor en el estudio de los métodos numéricos. En esencia, la *serie de Taylor* proporciona un medio para predecir el valor de una función en un punto en términos del valor de la función y sus derivadas en otro punto. En particular, el teorema establece que cualquier función suave puede aproximarse por un polinomio.

Una buena manera de comprender la serie de Taylor consiste en construirla término por término. Por ejemplo, el primer término de la serie es:

$$f(x_{i+1}) \cong f(x_i) \tag{4.2}$$

Esta relación, llamada la *aproximación de orden cero*, indica que el valor de f en el nuevo punto es el mismo que su valor en el punto anterior. Tal resultado tiene un sentido intuitivo, ya que si x_i y x_{i+1} están muy próximas entre sí, entonces es muy probable que el nuevo valor sea similar al anterior.

La ecuación (4.2) ofrece una estimación perfecta si la función que se va a aproximar es, de hecho, una constante. Sin embargo, si la función cambia en el intervalo, entonces se requieren los términos adicionales de la serie de Taylor, para obtener una mejor aproximación. Por ejemplo, la *aproximación de primer orden* se obtiene sumando otro término para obtener:

$$f(x_{i+1}) \cong f(x_i) + f'(x_i)(x_{i+1} - x_i) \tag{4.3}$$

El término adicional de primer orden consiste en una pendiente $f'(x_i)$ multiplicada por la distancia entre x_i y x_{i+1}. Por lo tanto, la expresión representa ahora una línea recta y es posible predecir un incremento o un decremento de la función entre x_i y x_{i+1}.

Aunque la ecuación (4.3) puede predecir un cambio, sólo es exacta para una línea recta o una tendencia *lineal*. Por lo tanto, se le agrega a la serie un término de *segundo orden* para obtener algo de la curvatura, que pudiera presentar la función:

$$f(x_{i+1}) \cong f(x_i) + f'(x_i)(x_{i+1} - x_i) + \frac{f''(x_i)}{2!}(x_{i+1} - x_i)^2 \tag{4.4}$$

De forma similar, se agregan términos adicionales para desarrollar la expansión completa de la serie de Taylor:

Cuadro 4.1 Teorema de Taylor

Teorema de Taylor

Si la función f y sus primeras $n + 1$ derivadas son continuas en un intervalo que contiene a y x, entonces el valor de la función en x está dado por

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n \tag{C4.1.1}$$

donde el residuo R_n se define como

$$R_n = \int_a^x \frac{(x-t)^n}{n!} f^{(n+1)}(t)dt \tag{C4.1.2}$$

donde t es una variable muda. La ecuación (C4.1.1) se llama *serie de Taylor* o *fórmula de Taylor.* Si se omite el residuo, el lado derecho de la ecuación (C4.1.1) es la aproximación del polinomio de Taylor para $f(x)$. En esencia, el teorema establece que cualquier función suave puede aproximarse mediante un polinomio.

La ecuación (C4.1.2) es sólo una manera, denominada la *forma integral*, mediante la cual puede expresarse el residuo. Se obtiene una formulación alternativa basándose en el teorema del valor medio para integrales.

Primer teorema del valor medio para integrales

Si la función g es continua e integrable en un intervalo que contenga a y x, entonces existe un punto ξ entre a y x tal que

$$\int_a^x g(t)dt = g(\xi)(x-a) \tag{C4.1.3}$$

En otras palabras, el teorema establece que la integral puede representarse por un valor promedio de la función $g(\xi)$ multiplicado por la longitud del intervalo $x - a$. Como el promedio debe encontrarse entre los valores mínimo y máximo del intervalo, existe un punto $x = \xi$ en el cual la función toma el valor promedio.

El primer teorema es, de hecho, un caso especial del segundo teorema del valor medio para integrales.

Segundo teorema del valor medio para integrales

Si las funciones g y h son continuas e integrables en un intervalo que contiene a y x, y h no cambia de signo en el intervalo, entonces existe un punto ξ entre a y x tal que

$$\int_a^x g(t)h(t)dt = g(\xi)\int_a^x h(t)dt \tag{C4.1.4}$$

La ecuación (C4.1.3) es equivalente a la ecuación (C4.1.4) con $h(t) = 1$.

El segundo teorema se aplica a la ecuación (C4.1.2) con

$$g(t) = f^{(n+1)}(t) \qquad h(t) = \frac{(x-t)^n}{n!}$$

Conforme t varía de a a x, $h(t)$ es continua y no cambia de signo. Por lo tanto, si $f^{(n+1)}(t)$ es continua, entonces se satisface el teorema del valor medio para integrales y

$$R_n = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}$$

Esta ecuación es conocida como la *forma de Lagrange* o *derivada* del residuo.

$$f(x_{i+1}) = f(x_i) + f'(x_i)(x_{i+1}-x_i) + \frac{f''(x_i)}{2!}(x_{i+1}-x_i)^2 + \frac{f^{(3)}(x_i)}{3!}(x_{i+1}-x_i)^3 + \cdots + \frac{f^{(n)}(x_i)}{n!}(x_{i+1}-x_i)^n + R_n \tag{4.5}$$

Observe que debido a que la ecuación (4.5) es una serie infinita, el signo igual reemplaza al signo de aproximación que se utiliza en las ecuaciones (4.2) a (4.4). Se incluye un término residual para considerar todos los términos desde el $n + 1$ hasta infinito:

$$R_n = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x_{i+1}-x_i)^{n+1} \tag{4.6}$$

donde el subíndice n indica que éste es el residuo de la aproximación de n-ésimo orden y ξ es un valor de x que se encuentra en algún punto entre x_i y x_{i+1}. La ξ es tan importante que se dedica una sección completa (sección 4.1.1) para su estudio. Por ahora es suficiente darse cuenta de que existe este valor que da una evaluación exacta del error.

Con frecuencia es conveniente simplificar la serie de Taylor definiendo un tamaño de paso o incremento $h = x_{i+1} - x_i$ y expresando la ecuación (4.5) como:

$$f(x_{i+1}) = f(x_i) + f'(x_i)h + \frac{f''(x_i)}{2!}h^2 + \frac{f^{(3)}(x_i)}{3!}h^3 + \cdots + \frac{f^{(n)}(x_i)}{n!}h^n + R_n \tag{4.7}$$

donde el término residual es ahora

$$R_n = \frac{f^{(n+1)}(\xi)}{(n+1)!} h^{n+1} \tag{4.8}$$

EJEMPLO 4.1 Aproximaciones de un polinomio mediante la serie de Taylor

Planteamiento del problema Use expansiones de la serie de Taylor de los órdenes cero hasta cuatro para aproximar la función

$$f(x) = -0.1x^4 - 0.15x^3 - 0.5x^2 - 0.25x + 1.2$$

desde $x_i = 0$ con $h = 1$. Esto es, estime el valor de la función en $x_{i+1} = 1$.

Solución Ya que se trata de una función conocida, es posible calcular valores de $f(x)$ entre 0 y 1. Los resultados (véase figura 4.1) indican que la función empieza en $f(0) = 1.2$ y hace una curva hacia abajo hasta $f(1) = 0.2$. Por lo tanto, el valor exacto que se intenta predecir es 0.2.

La aproximación de la serie de Taylor con $n = 0$ es [ecuación (4.2)]

$$f(x_{i+1}) \simeq 1.2$$

Como se muestra en la figura 4.1, la aproximación de orden cero es una constante. Usando esta formulación resulta un error de truncamiento [recuerde la ecuación (3.2)] de

$$E_t = 0.2 - 1.2 = -1.0$$

en $x = 1$.

Para $n = 1$, se debe determinar y evaluar la primera derivada en $x = 0$:

$$f'(0) = -0.4(0.0)^3 - 0.45(0.0)^2 - 1.0(0.0) - 0.25 = -0.25$$

La aproximación de primer orden es entonces [véase ecuación (4.3)]

$$f(x_{i+1}) \simeq 1.2 - 0.25h$$

que se emplea para calcular $f(1) = 0.95$. La aproximación empieza a coincidir con la trayectoria hacia abajo de la función en forma de una línea recta inclinada (véase figura 4.1). De esta manera, el error de truncamiento se reduce a

$$E_t = 0.2 - 0.95 = -0.75$$

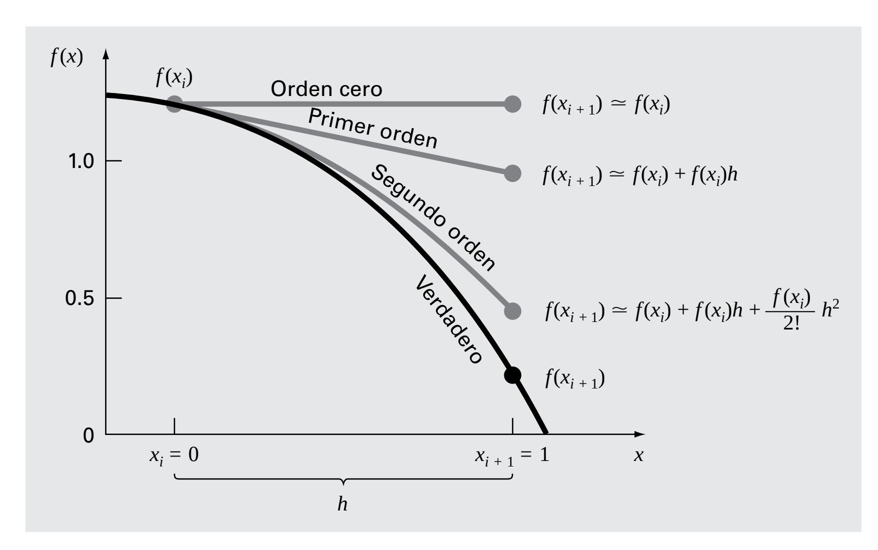


Figura 4.1
Aproximación de $f(x) = -0.1x^4 - 0.15x^3 - 0.5x^2 - 0.25x + 1.2$ en $x = 1$ mediante expansiones de la serie de Taylor de órdenes cero, primero y segundo.

Para $n = 2$, se evalúa la segunda derivada en $x = 0$:

$$f''(0) = -1.2(0.0)^2 - 0.9(0.0) - 1.0 = -1.0$$

Entonces, de acuerdo con la ecuación (4.4),

$$f(x_{i+1}) \simeq 1.2 - 0.25h - 0.5h^2$$

y sustituyendo $h = 1$, $f(1) = 0.45$. Al incluirse la segunda derivada se añade una curvatura descendente que proporciona una mejor estimación, como se muestra en la figura 4.1. Además, el error de truncamiento se reduce a $0.2 - 0.45 = -0.25$.

Los términos adicionales mejoran aún más la aproximación. En efecto, la inclusión de la tercera y de la cuarta derivadas da como resultado exactamente la misma ecuación del principio:

$$f(x) = 1.2 - 0.25h - 0.5h^2 - 0.15h^3 - 0.1h^4$$

donde el término residual es

$$R_4 = \frac{f^{(5)}(\xi)}{5!}h^5 = 0$$

ya que la quinta derivada de un polinomio de cuarto orden es cero. Por consiguiente, la expansión de la serie de Taylor hasta la cuarta derivada da una estimación exacta para $x_{i+1} = 1$:

$$f(1) = 1.2 - 0.25(1) - 0.5(1)^2 - 0.15(1)^3 - 0.1(1)^4 = 0.2$$

En general, la expansión de la serie de Taylor de n-ésimo orden será exacta para un polinomio de n-ésimo orden. Para otras funciones continuas y diferenciables, como las exponenciales y las senoidales, no se obtiene una estimación exacta con un número finito de términos. Cada uno de los términos adicionales contribuye, aunque sea con poco, al mejoramiento de la aproximación. Esto se muestra en el ejemplo 4.2, donde se obtendría un resultado exacto únicamente si se le agrega un número infinito de términos.

Aunque lo anterior es cierto, el valor práctico de las expansiones de la serie de Taylor estriba, en la mayoría de los casos, en que el uso de pocos términos que darán una aproximación lo suficientemente cercana a la solución verdadera para propósitos prácticos. La determinación de cuántos términos se requieren para obtener una "aproximación razonable" se basa en el término residual de la expansión. Recuerde que el término residual es de la forma general de la ecuación (4.8). Dicha fórmula tiene dos grandes inconvenientes. Primero, ξ no se conoce con exactitud, sino que sólo se sabe que está entre x_i y x_{i+1}. Segundo, para la evaluación de la ecuación (4.8) se requiere determinar la $(n + 1)$-ésima derivada de $f(x)$. Para hacerlo, se necesita conocer $f(x)$. Pero si ya se conoce $f(x)$, entonces no hay razón para realizar la expansión de la serie de Taylor.

A pesar de este dilema, la ecuación (4.8) aún resulta útil para la evaluación de errores de truncamiento. Esto se debe a que se *tiene* control sobre el término h de la ecuación. En otras palabras, es posible decidir qué tan lejos de x se desea evaluar $f(x)$ y controlar el número de términos que queremos tener en la expansión. Por esto, la ecuación (4.8) se expresa usualmente como

$$R_n = O(h^{n+1})$$

donde la nomenclatura $O(h^{n+1})$ significa que el error de truncamiento es de orden h^{n+1}. Es decir, el error es proporcional al incremento h elevado a la $(n + 1)$-ésima potencia. Aunque esta aproximación no implica nada en relación con la magnitud de las derivadas que multiplican h^{n+1}, es extremadamente útil para evaluar el error comparativo de los métodos numéricos que se basan en expansiones de la serie de Taylor. Por ejemplo, si el error es $O(h)$ y el incremento se reduce a la mitad, entonces el error también se reducirá a la mitad. Por otro lado, si el error es $O(h^2)$ y el incremento se reduce a la mitad, entonces el error se reducirá a una cuarta parte.

En general, se considera que el error de truncamiento disminuye agregando términos a la serie de Taylor. En muchos casos, si h es suficientemente pequeño, entonces el término de primer orden

y otros términos de orden inferior causan un porcentaje desproporcionadamente alto del error. Esta propiedad se ilustra en el ejemplo siguiente.

EJEMPLO 4.2 Uso de la expansión de la serie de Taylor para aproximar una función con un número infinito de derivadas

Planteamiento del problema Utilice expansiones de la serie de Taylor con n desde 0 hasta 6 para aproximar $f(x) = \cos x$ en $x_{i+1} = \pi/3$ con base en el valor de $f(x)$ y sus derivadas en $x_i = \pi/4$. Observe que esto significa que $h = \pi/3 - \pi/4 = \pi/12$.

Solución Como en el ejemplo 4.1, el conocimiento de la función original implica que se puede determinar el valor exacto de $f(\pi/3) = 0.5$.

La aproximación de orden cero es [ecuación (4.3)]

$$f\left(\frac{\pi}{3}\right) \cong \cos\left(\frac{\pi}{4}\right) = 0.707106781$$

que representa un error relativo porcentual de

$$\varepsilon_t = \frac{0.5 - 0.707106781}{0.5} 100\% = -41.4\%$$

Para la aproximación de primer orden, se agrega el término de la primera derivada donde $f'(x) = -\text{sen } x$:

$$f\left(\frac{\pi}{3}\right) \cong \cos\left(\frac{\pi}{4}\right) - \text{sen}\left(\frac{\pi}{4}\right)\left(\frac{\pi}{12}\right) = 0.521986659$$

que tiene $\varepsilon_t = -4.40$ por ciento.

Para la aproximación de segundo orden, se agrega el término de la segunda derivada donde $f''(x) = -\cos x$:

$$f\left(\frac{\pi}{3}\right) \cong \cos\left(\frac{\pi}{4}\right) - \text{sen}\left(\frac{\pi}{4}\right)\left(\frac{\pi}{12}\right) - \frac{\cos(\pi/4)}{2}\left(\frac{\pi}{12}\right)^2 = 0.497754491$$

con $\varepsilon_t = 0.449$ por ciento. Entonces, al agregar más términos a la serie se obtiene una mejor aproximación.

Este proceso continúa y sus resultados se enlistan, como en la tabla 4.1. Observe que las derivadas nunca se aproximan a cero, como es el caso con el polinomio del ejemplo 4.1. Por lo tanto, cada término que se le agrega a la serie genera una mejor aproximación. Sin embargo, observe también que la mejor aproximación se consigue con los primeros términos. En este caso, al agregar el tercer término, el error se redujo al 2.62×10^{-2}%, lo cual significa que se alcanzó el 99.9738% del valor exacto. Por consiguiente, aunque la adición de más términos reducirá el error, la mejoría será mínima.

Tabla 4.1 Aproximaciones mediante la serie de Taylor de $f(x) = \cos x$ en $x_{i+1} = \pi/3$ usando como punto base $\pi/4$. Los valores se presentan para varios órdenes (n) de aproximación.

Orden n	$f^{(n)}(x)$	$f(\pi/3)$	ε_t
0	cos x	0.707106781	–41.4
1	–sen x	0.521986659	–4.4
2	–cos x	0.497754491	0.449
3	sen x	0.499869147	2.62×10^{-2}
4	cos x	0.500007551	-1.51×10^{-3}
5	–sen x	0.500000304	-6.08×10^{-5}
6	–cos x	0.499999988	2.40×10^{-6}

4.1.1 El residuo en la expansión de la serie de Taylor

Antes de mostrar cómo se utiliza la serie de Taylor en la estimación de errores numéricos, se debe explicar por qué se incluye el argumento ξ en la ecuación (4.8). El desarrollo matemático se presenta en el cuadro 4.1. Ahora se expondrá una interpretación más visual. Después se extiende este caso específico a una formulación más general.

Suponga que se trunca la expansión de la serie de Taylor [ecuación (4.7)] después del término de orden cero para obtener:

$$f(x_{i+1}) \cong f(x_i)$$

En la figura 4.2 se muestra una representación gráfica de esta predicción de orden cero. El residuo o error de esta predicción, que se indica también en la figura, consiste de la serie infinita de términos que fueron truncados:

$$R_0 = f'(x_i)h + \frac{f''(x_i)}{2!}h^2 + \frac{f^{(3)}(x_i)}{3!}h^3 + \cdots$$

Obviamente no resulta conveniente manipular el residuo en este formato de serie infinita. Se obtiene una simplificación truncando el residuo mismo de la siguiente manera

$$R_0 \cong f'(x_i)h \tag{4.9}$$

Aunque como se mencionó en la sección previa, por lo común las derivadas de orden inferior cuentan mucho más en el residuo que los términos de las derivadas de orden superior; este resultado todavía es inexacto, ya que se han despreciado los términos de segundo orden y de órdenes superiores. Esta "inexactitud" se denota mediante el símbolo de aproximación a la igualdad ($\cong$) empleado en la ecuación (4.9).

Una simplificación alternativa que transforma la aproximación en una equivalencia está basada en un esquema gráfico. Como se muestra en la figura 4.3 el *teorema del valor medio para la derivada* establece que si una función $f(x)$ y su primera derivada son continuas en el intervalo de x_i a x_{i+1}, entonces existe al menos un punto en la función que tiene una pendiente, denotada por $f'(\xi)$, que es paralela a la línea que une $f(x_i)$ y $f(x_{i+1})$. El parámetro ξ marca el valor x donde se presenta la pendiente (figura 4.3). Una ilustración física de este teorema es la siguiente: si usted viaja entre dos puntos a una velocidad promedio, habrá al menos un momento durante el curso del viaje en que usted se mueve a esa velocidad promedio.

Al utilizar este teorema resulta fácil darse cuenta, como se muestra en la figura 4.3, de que la pendiente $f'(\xi)$ es igual al cociente de la elevación R_0 entre el recorrido h, o

$$f'(\xi) = \frac{R_0}{h}$$

que se puede reordenar para obtener

$$R_0 = f'(\xi)h \tag{4.10}$$

Por lo tanto, se ha obtenido la versión de orden cero de la ecuación (4.8). Las versiones de orden superior son tan sólo una extensión lógica del razonamiento usado para encontrar la ecuación (4.10). La versión de primer orden es

$$R_1 = \frac{f''(\xi)}{2!}h^2 \tag{4.11}$$

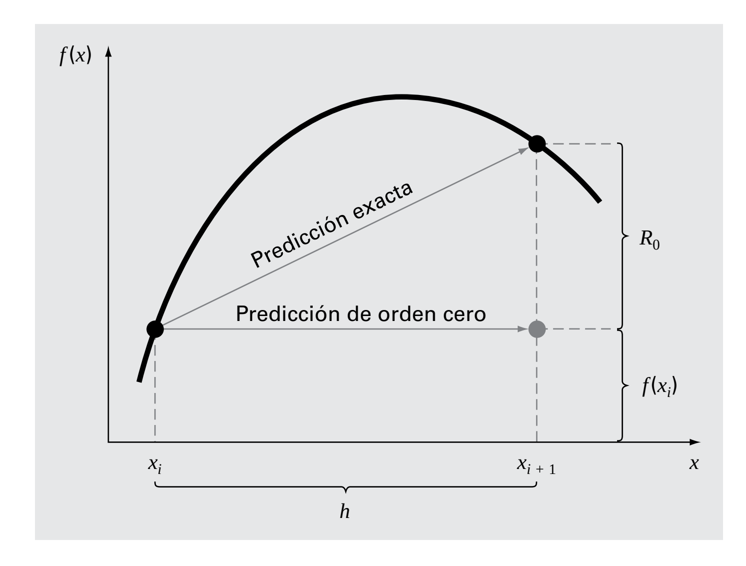


Figura 4.2
Representación gráfica de una predicción de orden cero con la serie de Taylor y del residuo.

En este caso, el valor de ξ será el valor de x que corresponde a la derivada de segundo orden que hace exacta a la ecuación (4.11). Es posible obtener versiones similares de orden superior a partir de la ecuación (4.8).

4.1.2 Uso de la serie de Taylor para estimar los errores de truncamiento

Aunque la serie de Taylor será muy útil en la estimación de los errores de truncamiento a lo largo de este libro, quizá no resulte claro cómo la expansión se aplica a los métodos numéricos. De hecho, esto ya se hizo en el ejemplo de la caída del paracaidista. Recuerde que el objetivo de los ejemplos 1.1 y 1.2 fue predecir la velocidad como una función del tiempo. Es decir, se deseaba determinar $v(t)$. Como se especificó en la ecuación (4.5), $v(t)$ se puede expandir en una serie de Taylor del siguiente modo:

$$v(t_{i+1}) = v(t_i) + v'(t_i)(t_{i+1} - t_i) + \frac{v''(t_i)}{2!}(t_{i+1} - t_i)^2 + \cdots + R_n \qquad (4.12)$$

Ahora, truncando la serie después del término con la primera derivada, se obtiene:

$$v(t_{i+1}) = v(t_i) + v'(t_i)(t_{i+1} - t_i) + R_1 \qquad (4.13)$$

En la ecuación (4.13) se despeja obteniendo

$$v'(t_i) = \underbrace{\frac{v(t_{i+1}) - v(t_i)}{t_{i+1} - t_i}}_{\substack{\text{Aproximación}\\ \text{de primer orden}}} - \underbrace{\frac{R_1}{t_{i+1} - t_1}}_{\substack{\text{Error de}\\ \text{truncamiento}}} \qquad (4.14)$$

La primera parte de la ecuación (4.14) es exactamente la misma relación que se usó para aproximar la derivada del ejemplo 1.2 [ecuación (1.11)]. Sin embargo, con el método de la serie de Taylor se ha obtenido una estimación del error de truncamiento asociado con esta aproximación de la derivada. Utilizando las ecuaciones (4.6) y (4.14) se tiene

$$\frac{R_1}{t_{i+1} - t_i} = \frac{v''(\xi)}{2!}(t_{i+1} - t_i) \qquad (4.15)$$

o

$$\frac{R_1}{t_{i+1} - t_i} = O(t_{i+1} - t_i) \qquad (4.16)$$

Por lo tanto, la estimación de la derivada [ecuación (1.11) o la primera parte de la ecuación (4.14)] tiene un error de truncamiento de orden $t_{i+1} - t_i$. En otras palabras, el error en nuestra aproximación de la derivada debería ser proporcional al tamaño del incremento. Entonces, si éste se divide a la mitad, se esperaría que el error de la derivada se reduzca a la mitad.

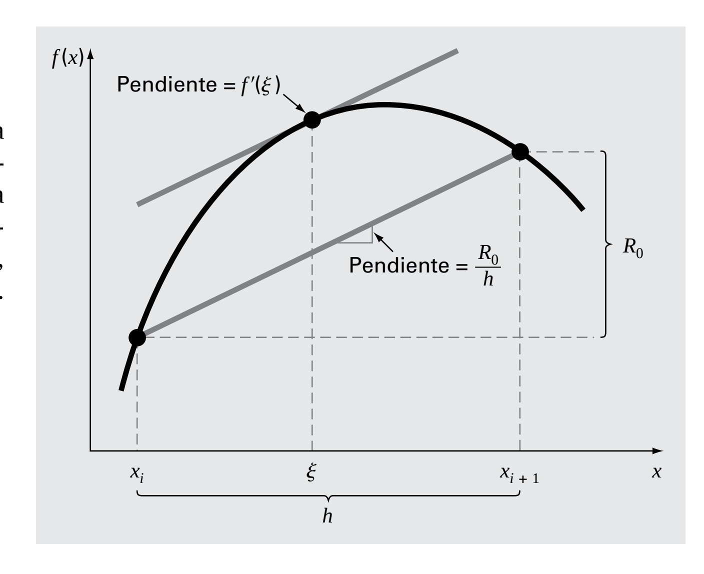


Figura 4.3
Representación gráfica del teorema del valor medio para la derivada.

EJEMPLO 4.3 El efecto de no linealidad y del tamaño del incremento en la aproximación de la serie de Taylor

Planteamiento del problema En la figura 4.4 se grafica la función

$$f(x) = x^m \tag{E4.3.1}$$

para $m = 1, 2, 3$ y 4 en el rango de $x = 1$ a 2. Observe que para $m = 1$ la función es lineal, y conforme *m* se incrementa, se presenta mayor curvatura o no linealidad dentro de la función.

Utilizar la serie de Taylor de primer orden para aproximar la función con diversos valores del exponente *m* y del tamaño de incremento *h*.

Solución La ecuación (E4.3.1) se aproxima por una expansión de la serie de Taylor de primer orden:

$$f(x_{i+1}) = f(x_i) + mx_i^{m-1}h \tag{E4.3.2}$$

la cual tiene un residuo de

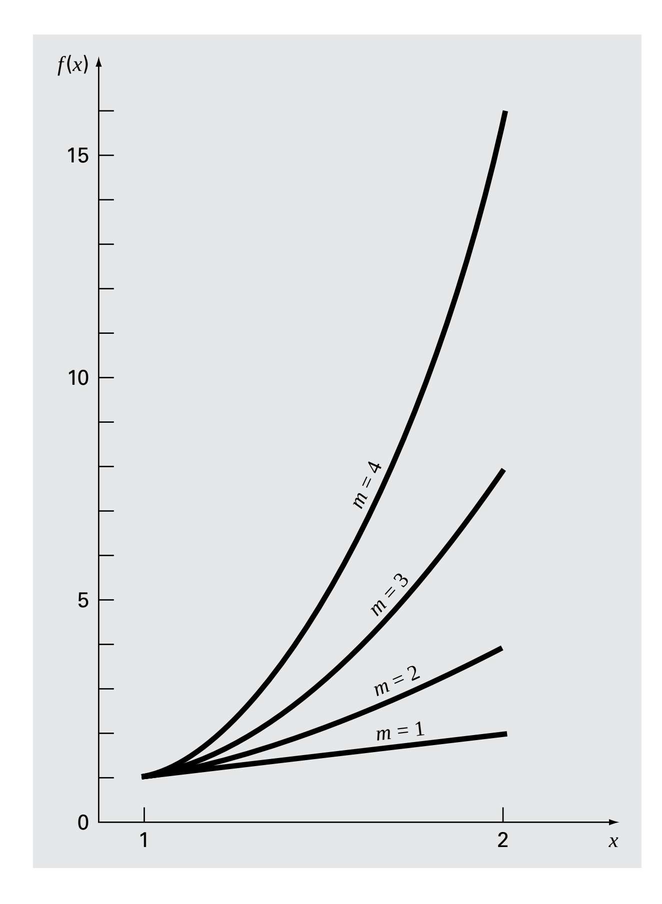
$$R_1 = \frac{f''(x_i)}{2!}h^2 + \frac{f^{(3)}(x_i)}{3!}h^3 + \frac{f^{(4)}(x_i)}{4!}h^4 + \cdots$$

Primero, puede examinarse cómo se comporta la aproximación conforme *m* aumenta; es decir, conforme la función se vuelve más no lineal. Para $m = 1$, el valor exacto de la función en $x = 2$ es 2. La serie de Taylor nos da

$$f(2) = 1 + 1(1) = 2$$

y

$$R_1 = 0$$



El residuo es cero porque la segunda derivada y las derivadas de orden superior de una función lineal son cero. Entonces, como es de esperarse, la expansión de la serie de Taylor de primer orden es perfecta cuando la función de que se trata es lineal.

Para $m = 2$, el valor real es $f(2) = 2^2 = 4$. La aproximación de la serie de Taylor de primer orden es

$$f(2) = 1 + 2(1) = 3$$

y

$$R_1 = \tfrac{2}{2}(1)^2 + 0 + 0 + \cdots = 1$$

Debido a que la función es una parábola, la aproximación mediante una línea recta da por resultado una discrepancia. Observe que el residuo se determina en forma exacta.

Figura 4.4
Gráfica de la función $f(x) = x^m$ para $m = 1, 2, 3$ y 4. Note que la función tiende a ser más no lineal cuando aumenta *m*.

Para $m = 3$, el valor real es $f(2) = 2^3 = 8$. La aproximación con la serie de Taylor es

$$f(2) = 1 + 3(1)^2(1) = 4$$

y

$$R_1 = \tfrac{6}{2}(1)^2 + \tfrac{6}{6}(1)^3 + 0 + 0 + \cdots = 4$$

Otra vez, hay una discrepancia que se puede determinar exactamente a partir de la serie de Taylor.

Para $m = 4$, el valor real es $f(2) = 2^4 = 16$. La aproximación con la serie de Taylor es

$$f(2) = 1 + 4(1)^3(1) = 5$$

y

$$R_1 = \tfrac{12}{2}(1)^2 + \tfrac{24}{6}(1)^3 + \tfrac{24}{24}(1)^4 + 0 + 0 + \cdots = 11$$

Considerando estos cuatro casos, se observa que R_1 se incrementa conforme la función empieza a ser cada vez más no lineal. Además, R_1 da cuenta exacta de la discrepancia, porque la ecuación (E4.3.1) es un simple monomio con un número finito de derivadas. Esto permite una completa determinación del residuo de la serie de Taylor.

Ahora examinemos la ecuación (E4.3.2) para el caso en que $m = 4$ y observe cómo R_1 cambia cuando el tamaño del incremento h varía. Para $m = 4$, la ecuación (E4.3.2) es

$$f(x+h) = f(x) + 4x_i^3 h$$

Si $x = 1$, $f(1) = 1$ entonces esta ecuación se expresa como

$$f(1 + h) = 1 + 4h$$

con un residuo de

$$R_1 = 6h^2 + 4h^3 + h^4$$

Lo cual nos lleva a la conclusión de que la discrepancia disminuirá conforme h se reduzca. Entonces, para valores suficientemente pequeños de h, el error debería ser proporcional a h^2. Es decir, conforme h se reduce a la mitad, el error se reduce a la cuarta parte. Este comportamiento se confirma en la tabla 4.2 y en la figura 4.5.

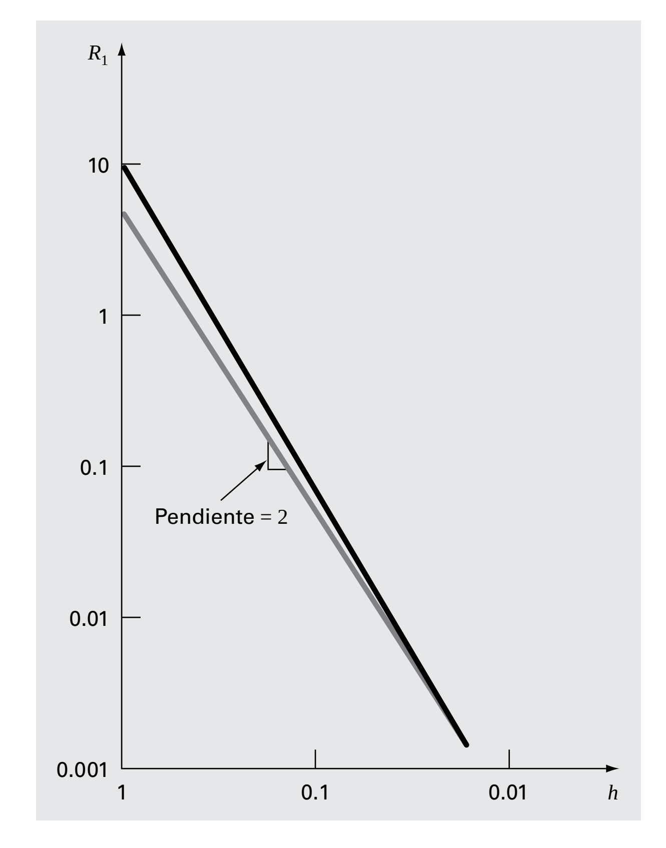
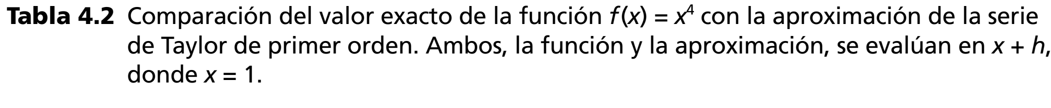


Figura 4.5
Gráfica en escala log-log del residuo R_1 para la aproximación de la función $f(x) = x^4$ mediante la serie de Taylor de primer orden contra el tamaño del incremento h. La línea con la pendiente 2 también se muestra para indicar que conforme h disminuye, el error se vuelve proporcional a h^2.


Tabla 4.2 Comparación del valor exacto de la función $f(x) = x^4$ con la aproximación de la serie de Taylor de primer orden. Ambos, la función y la aproximación, se evalúan en $x + h$, donde $x = 1$.

h	Exacto	Aproximación de primer orden	R_1
1	16	5	11
0.5	5.0625	3	2.0625
0.25	2.441406	2	0.441406
0.125	1.601807	1.5	0.101807
0.0625	1.274429	1.25	0.024429
0.03125	1.130982	1.125	0.005982
0.015625	1.063980	1.0625	0.001480

De esta forma, se concluye que el error de la aproximación por serie de Taylor de primer orden disminuye conforme *m* se aproxima a 1 y conforme *h* disminuye. Intuitivamente, esto significa que la serie de Taylor adquiere más exactitud cuando la función que se está aproximando se vuelve más semejante a una línea recta sobre el intervalo de interés. Esto se logra reduciendo el tamaño del intervalo o "enderezando" la función por reducción de *m*. Es obvio que dicha opción usualmente no está disponible en el mundo real porque las funciones para analizar son, en forma general, dictadas en el contexto del problema físico. En consecuencia, no se tiene control sobre la falta de linealidad y el único recurso consiste en reducir el tamaño del incremento o incluir términos adicionales de la expansión de la serie de Taylor.

4.1.3 Diferenciación numérica

A la ecuación (4.14) se le conoce con un nombre especial en el análisis numérico: *diferencias finitas divididas* y generalmente se representa como

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} + O(x_{i+1} - x_i) \tag{4.17}$$

o

$$f'(x_i) = \frac{\Delta f_i}{h} + O(h) \tag{4.18}$$

donde a Δf_i se le conoce como la *primera diferencia hacia adelante* y a *h* se le llama el tamaño del paso o incremento; esto es, la longitud del intervalo sobre el cual se realiza la aproximación. Se le llama diferencia "hacia adelante", porque usa los datos en *i* e *i* + 1 para estimar la derivada (figura 4.6*a*). Al término completo $\Delta f/h$ se le conoce como *primera diferencia finita dividida*.

Esta diferencia dividida hacia adelante es sólo una de tantas que pueden desarrollarse a partir de la serie de Taylor para la aproximación de derivadas numéricas. Por ejemplo, las aproximaciones de la primera derivada utilizando *diferencias hacia atrás* o *diferencias centradas* se pueden desarrollar de una manera similar a la de la ecuación (4.14). Las primeras usan valores en x_{i-1} y x_i (figura 4.6*b*); mientras que las segundas utilizan valores igualmente espaciados alrededor del punto donde la derivada está estimada (figura 4.6*c*). Es posible desarrollar aproximaciones más exactas de la primera derivada incluyendo términos de orden más alto de la serie de Taylor. Finalmente, todas las versiones anteriores se pueden desarrollar para derivadas de segundo orden, de tercer orden y de órdenes superiores. En las siguientes secciones se dan resúmenes breves que ilustran cómo se obtienen algunos de estos casos.

Aproximación a la primera derivada con diferencias hacia atrás La serie de Taylor se expande hacia atrás para calcular un valor anterior sobre la base del valor actual,

$$f(x_{i-1}) = f(x_i) - f'(x_i)h + \frac{f''(x_i)}{2!}h^2 - \cdots \tag{4.19}$$

Truncando la ecuación después de la primera derivada y reordenando los términos se obtiene

$$f'(x_i) \cong \frac{f(x_i) - f(x_{i-1})}{h} = \frac{\nabla f_1}{h} \tag{4.20}$$

donde el error es $O(h)$, y a ∇f_i se le conoce como *primera diferencia dividida hacia atrás*. Véase la figura 4.6*b* para una representación gráfica.

Aproximación a la primera derivada con diferencias centradas Una tercera forma de aproximar la primera derivada consiste en restar la ecuación (4.19) de la expansión de la serie de Taylor hacia adelante:

$$f(x_{i+1}) = f(x_i) + f'(x_i)h + \frac{f''(x_i)}{2!}h^2 + \cdots \tag{4.21}$$

para obtener

$$f(x_{i+1}) = f(x_{i-1}) + 2f'(x_i)h + \frac{2f^{(3)}(x_i)}{3!}h^3 + \cdots$$

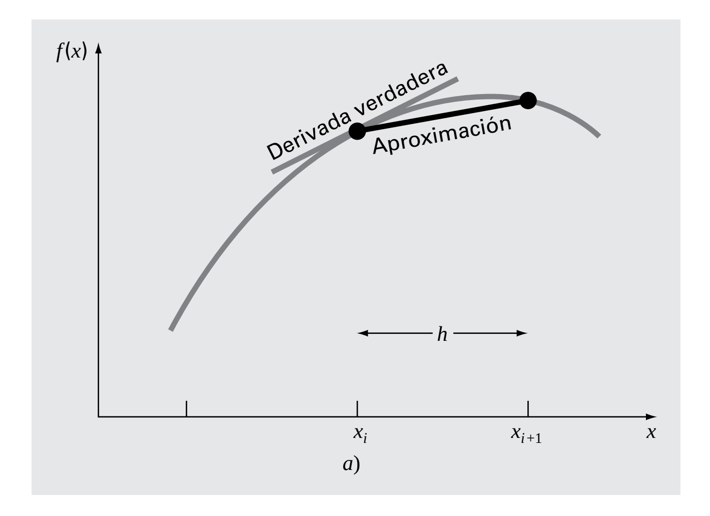
de donde se despeja

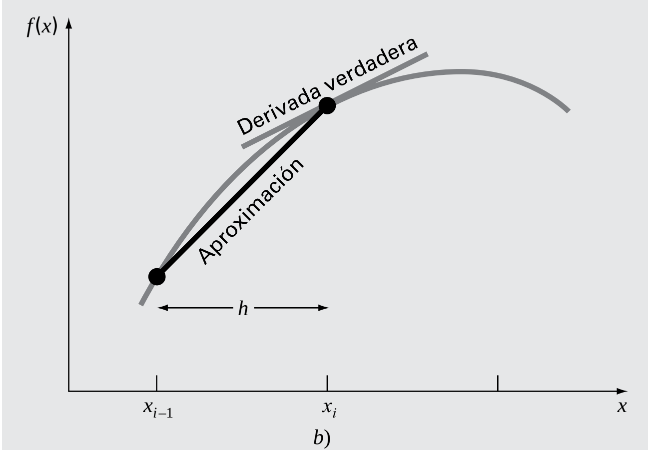
$$f'(x_i) = \frac{f(x_{i+1}) - f(x_{i-1})}{2h} - \frac{f^{(3)}(x_i)}{6}h^2 - \cdots$$

o

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_{i-1})}{2h} - O(h^2) \tag{4.22}$$

La ecuación (4.22) es una representación de las *diferencias centradas* de la primera derivada. Observe que el error de truncamiento es del orden de h^2 en contraste con las aproximaciones hacia adelante y hacia atrás, que fueron del orden de h. Por lo tanto, el análisis de la serie de Taylor ofrece la información práctica de que las diferencias centradas son una representación más exacta de la derivada (figura 4.6*c*). Por ejemplo, si disminuimos el tamaño del incremento a la mitad, usando diferencias hacia atrás o hacia adelante, el error de truncamiento se reducirá aproximadamente a la mitad; mientras que con diferencias centradas el error se reduciría a la cuarta parte.





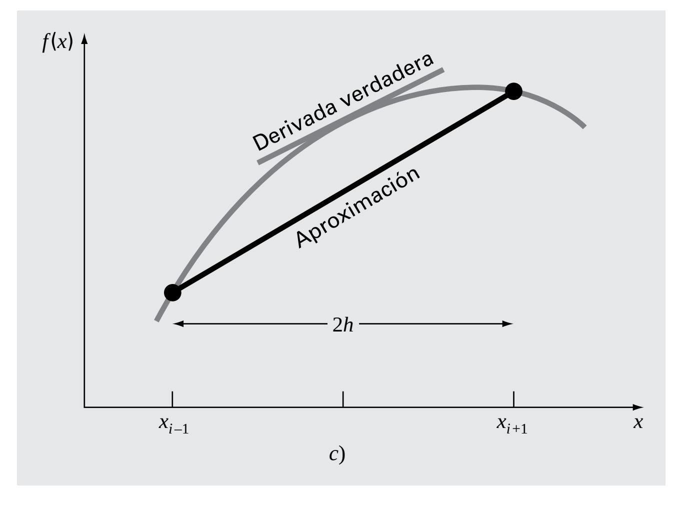


Figura 4.6
Gráfica de aproximaciones con diferencias finitas divididas de la primera derivada:*a*) hacia adelante, *b*) hacia atrás, *c*) centrales.

EJEMPLO 4.4 Aproximación de derivadas por diferencias finitas divididas

Planteamiento del problema Use aproximaciones con diferencias finitas hacia adelante y hacia atrás de $O(h)$ y una aproximación de diferencias centradas de $O(h^2)$ para estimar la primera derivada de

$$f(x) = -0.1x^4 - 0.15x^3 - 0.5x^2 - 0.25x + 1.25$$

en $x = 0.5$ utilizando un incremento de $h = 0.5$. Repita el cálculo con $h = 0.25$. Observe que la derivada se calcula directamente como

$$f'(x) = -0.4x^3 - 0.45x^2 - 1.0x - 0.25$$

y se puede utilizar para calcular el valor verdadero como $f'(0.5) = -0.9125$.

Solución Para $h = 0.5$, la función se emplea para determinar

$x_{i-1} = 0$ $\quad f(x_{i-1}) = 1.2$
$x_i = 0.5$ $\quad f(x_i) = 0.925$
$x_{i+1} = 1.0$ $\quad f(x_{i+1}) = 0.2$

Esos valores sirven para calcular las diferencias divididas hacia adelante [ecuación (4.17)],

$$f'(0.5) \cong \frac{0.2 - 0.925}{0.5} = -1.45 \qquad |\varepsilon_t| = 58.9\%$$

la diferencia dividida hacia atrás [ecuación (4.20)],

$$f'(0.5) \cong \frac{0.925 - 1.2}{0.5} = -0.55 \qquad |\varepsilon_t| = 39.7\%$$

y la diferencia dividida centrada [ecuación (4.22)],

$$f'(0.5) \cong \frac{0.2 - 1.2}{1.0} = -1.0 \qquad |\varepsilon_t| = 9.6\%$$

Para $h = 0.25$,

$x_{i-1} = 0.25$ $\quad f(x_{i-1}) = 1.10351563$
$x_i = 0.5$ $\quad f(x_i) = 0.925$
$x_{i+1} = 0.75$ $\quad f(x_{i+1}) = 0.63632813$

que se utilizan para calcular la diferencia dividida hacia adelante,

$$f'(0.5) \cong \frac{0.63632813 - 0.925}{0.25} = -1.155 \qquad |\varepsilon_t| = 26.5\%$$

la diferencia dividida hacia atrás,

$$f'(0.5) \cong \frac{0.925 - 1.10351563}{0.25} = -0.714 \qquad |\varepsilon_t| = 21.7\%$$

y la diferencia dividida centrada,

$$f'(0.5) \cong \frac{0.63632813 - 1.10351563}{0.5} = -0.934 \qquad |\varepsilon_t| = 2.4\%$$

Para ambos tamaños de paso, la aproximación en diferencias centrales es más exacta que las diferencias hacia adelante y hacia atrás. También, como se pronosticó con el análisis de la serie de Taylor, dividiendo a la mitad el incremento, se tiene aproximadamente la mitad del error en las diferencias hacia atrás y hacia adelante y una cuarta parte de error en las diferencias centradas.

Aproximaciones por diferencias finitas para derivadas de orden superior Además de las primeras derivadas, la expansión en serie de Taylor sirve para obtener estimaciones numéricas de las derivadas de orden superior. Para esto, se escribe la expansión en serie de Taylor hacia adelante para $f(x_{i+2})$ en términos de $f(x_i)$:

$$f(x_{i+2}) = f(x_i) + f'(x_i)(2h) + \frac{f''(x_i)}{2!}(2h)^2 + \cdots \tag{4.23}$$

La ecuación (4.21) se multiplica por 2 y se resta de la ecuación (4.23) para obtener

$$f(x_{i+2}) - 2\,f(x_{i+1}) = -f(x_i) + f''(x_i)h^2 + \cdots$$

de donde se despeja

$$f''(x_i) = \frac{f(x_{i+2}) - 2\,f(x_{i+1}) + f(x_i)}{h^2} + O(h) \tag{4.24}$$

Esta relación se llama la *segunda diferencia finita dividida hacia adelante*. Manipulaciones similares se emplean para obtener la versión hacia atrás

$$f''(x_i) = \frac{f(x_i) - 2\,f(x_{i-1}) + f(x_{i-2})}{h^2} + O(h)$$

y la versión centrada

$$f''(x_i) = \frac{f(x_{i+1}) - 2\,f(x_i) + f(x_{i-1})}{h^2} + O(h^2)$$

Como fue el caso con las aproximaciones de la primera derivada, el caso centrado es más exacto. Observe también que la versión centrada puede ser expresada en forma alternativa como

$$f''(x_i) \cong \frac{\dfrac{f(x_{i+1}) - f(x_i)}{h} - \dfrac{f(x_i) - f(x_{i-1})}{h}}{h}$$

Así, como la segunda derivada es una derivada de la derivada, la aproximación de la segunda diferencia dividida es una diferencia de dos primeras diferencias divididas.

Se volverá al tema de la diferenciación numérica en el capítulo 23. Aquí hemos presentado este tema porque es un muy buen ejemplo de por qué la serie de Taylor es importante en los métodos numéricos. Además, varias de las fórmulas vistas en esta sección se emplearán antes del capítulo 23.

4.2 PROPAGACIÓN DEL ERROR

El propósito de esta sección consiste en estudiar cómo los errores en los números pueden propagarse a través de las funciones matemáticas. Por ejemplo, si se multiplican dos números que tienen errores, nos gustaría estimar el error de este producto.

4.2.1 Funciones de una sola variable

Suponga que se tiene la función $f(x)$ que es dependiente de una sola variable independiente x. Considere que $\tilde{x}$ es una aproximación de x. Por lo tanto, se desearía evaluar el efecto de la discrepancia entre x y $\tilde{x}$ en el valor de la función. Esto es, se desearía estimar

$$\Delta f(\tilde{x}) = |f(x) - f(\tilde{x})|$$

El problema para evaluar $\Delta f(\tilde{x})$ es que se desconoce $f(x)$ porque se desconoce x. Se supera esta dificultad si $\tilde{x}$ está cercana a x y $f(\tilde{x})$ es continua y diferenciable. Si se satisfacen estas condiciones se utiliza la serie de Taylor para calcular $f(x)$ cerca de $f(\tilde{x})$,

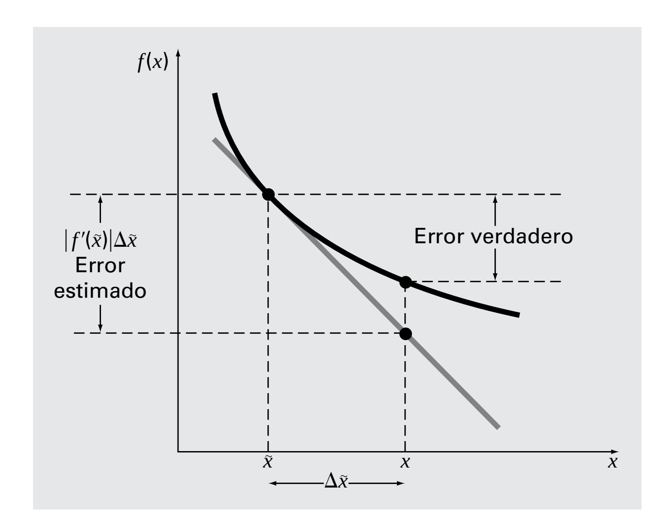


Figura 4.7
Representación gráfica de la propagación del error de primer orden.

$$f(x) = f(\tilde{x}) + f'(\tilde{x})(x - \tilde{x}) + \frac{f''(\tilde{x})}{2}(x - \tilde{x})^2 + \cdots$$

Quitando el segundo término, los de orden superior, y reordenando, se obtiene

$$f(x) - f(\tilde{x}) \cong f'(\tilde{x})(x - \tilde{x})$$

o

$$\Delta f(\tilde{x}) = |f'(\tilde{x})| \Delta\tilde{x} \tag{4.25}$$

donde $\Delta f(\tilde{x}) = |f(x) - f(\tilde{x})|$ representa una estimación del error de la función y $\Delta\tilde{x} = |x - \tilde{x}|$ representa una estimación del error de *x*. La ecuación (4.25) proporciona la capacidad de aproximar el error en $f(x)$ dando la derivada de una función y una estimación del error en la variable independiente. La figura 4.7 es una gráfica que representa esta operación.

EJEMPLO 4.5 Propagación del error en una función de una variable

Planteamiento del problema Dado un valor de $\tilde{x} = 2.5$ con un error $\Delta\tilde{x} = 0.01$, estime el error resultante en la función $f(x) = x^3$.

Solución Con la ecuación (4.25),

$$\Delta f(\tilde{x}) \cong 3(2.5)^2(0.01) = 0.1875$$

Ya que $f(2.5) = 15.625$, se pronostica que

$$f(2.5) = 15.625 \pm 0.1875$$

o que el valor verdadero se encuentra entre 15.4375 y 15.8125. De hecho, si *x* fuera realmente 2.49, la función se evaluaría como 15.4382, y si *x* fuera 2.51, el valor de la función sería 15.8132. Para este caso, el análisis del error de primer orden proporciona una estimación adecuada del error verdadero.

4.2.2 Funciones de más de una variable

El enfoque anterior puede generalizarse a funciones que sean dependientes de más de una variable independiente, lo cual se realiza con una versión para varias variables de la serie de Taylor. Por ejemplo, si se tiene una función de dos variables independientes, *u* y *v*, la serie de Taylor se escribe como

$$\begin{aligned} f(u_{i+1}, v_{i+1}) = f(u_i, v_i) &+ \frac{\partial f}{\partial u}(u_{i+1} - u_i) + \frac{\partial f}{\partial v}(v_{i+1} - v_i) \\ &+ \frac{1}{2!}\left[\frac{\partial^2 f}{\partial u^2}(u_{i+1} - u_i)^2 + 2\frac{\partial^2 f}{\partial u \partial v}(u_{i+1} - u_i)(v_{i+1} - v_i) \right. \\ &\left. + \frac{\partial^2 f}{\partial v^2}(v_{i+1} - v_i)^2\right] + \cdots \end{aligned} \tag{4.26}$$

donde todas las derivadas parciales se evalúan en el punto base *i*. Si no se consideran todos los términos de segundo orden y de orden superior, de la ecuación (4.26) puede despejarse

$$\Delta f(\tilde{u},\tilde{v}) = \left|\frac{\partial f}{\partial u}\right| \Delta\tilde{u} + \left|\frac{\partial f}{\partial v}\right| \Delta\tilde{v}$$

donde $\Delta\tilde{u}$ y $\Delta\tilde{v}$ son estimaciones del error en u y v, respectivamente.

Para n variables independientes $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n$ teniendo errores $\Delta\tilde{x}_1, \Delta\tilde{x}_2, \ldots, \Delta x_n$ se satisface la siguiente relación general:

$$\Delta f(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n) \cong \left|\frac{\partial f}{\partial x_1}\right| \Delta\tilde{x}_1 + \left|\frac{\partial f}{\partial x_2}\right| \Delta\tilde{x}_2 + \cdots + \left|\frac{\partial f}{\partial x_n}\right| \Delta\tilde{x}_n \tag{4.27}$$

EJEMPLO 4.6 Propagación del error en una función con varias variables

Planteamiento del problema La deflexión y de la punta de un mástil en un bote de vela es

$$y = \frac{FL^4}{8EI}$$

donde F = una carga lateral uniforme (N/m), L = altura (m), E = el módulo de elasticidad (N/m^2) e I = el momento de inercia (m^4). Estime el error en y con los siguientes datos:

$\tilde{F}$ = 750 N/m	$\Delta\tilde{F}$ = 30 N/m
$\tilde{L}$ = 9 m	$\Delta\tilde{L}$ = 0.03 m
$\tilde{E}$ = 7.5 × 10^9 N/m^2	$\Delta\tilde{E}$ = 5 × 10^7 N/m^2
$\tilde{I}$ = 0.0005 m^4	$\Delta\tilde{I}$ = 0.000005 m^4

Solución Empleando la ecuación (4.27) se tiene

$$\Delta y(\tilde{F}, \tilde{L}, \tilde{E}, \tilde{I}) = \left|\frac{\partial y}{\partial F}\right| \Delta\tilde{F} + \left|\frac{\partial y}{\partial L}\right| \Delta\tilde{L} + \left|\frac{\partial y}{\partial E}\right| \Delta\tilde{E} + \left|\frac{\partial y}{\partial I}\right| \Delta\tilde{I}$$

o

$$\Delta y(\tilde{F}, \tilde{L}, \tilde{E}, \tilde{I}) \cong \frac{\tilde{L}^4}{8\tilde{E}\tilde{I}} \Delta\tilde{F} + \frac{\tilde{F}\tilde{L}^3}{2\tilde{E}\tilde{I}} \Delta\tilde{L} + \frac{\tilde{F}\tilde{L}^4}{8\tilde{E}^2\tilde{I}} \Delta\tilde{E} + \frac{\tilde{F}\tilde{L}^4}{8\tilde{E}\tilde{I}^2} \Delta\tilde{I}$$

Sustituyendo los valores adecuados se obtiene

$$\Delta y = 0.006561 + 0.002187 + 0.001094 + 0.00164 = 0.011482$$

Por consiguiente, y = 0.164025 ± 0.011482. En otras palabras, y está entre 0.152543 y 0.175507 m. La validez de estos estimados se puede verificar sustituyendo los valores extremos de las variables en la ecuación para generar un mínimo exacto de

$$y_{\text{mín}} = \frac{720(8.97)^4}{8(7.55 \times 10^9)0.000505} = 0.152818$$

y

$$y_{\text{máx}} = \frac{780(9.03)^4}{8(7.45 \times 10^9)0.000495} = 0.175790$$

Así, los estimados de primer orden están razonablemente cercanos a los valores exactos.

La ecuación (4.27) se utiliza para definir relaciones en la propagación de errores con las operaciones matemáticas comunes. Los resultados se resumen en la tabla 4.3. Se deja el desarrollo de estas fórmulas como un ejercicio de tarea.

Tabla 4.3 El error estimado relacionado con las operaciones matemáticas comunes usando números inexactos $\tilde{u}$ y $\tilde{v}$.

Operación		Error estimado
Adición	$\Delta(\tilde{u} + \tilde{v})$	$\Delta\tilde{u} + \Delta\tilde{v}$
Sustracción	$\Delta(\tilde{u} - \tilde{v})$	$\Delta\tilde{u} + \Delta\tilde{v}$
Multiplicación	$\Delta(\tilde{u} \times \tilde{v})$	$\|\tilde{u}\|\Delta\tilde{v} + \|\tilde{v}\|\Delta\tilde{u}$
División	$\Delta\left(\frac{\tilde{u}}{\tilde{v}}\right)$	$\frac{\|\tilde{u}\|\Delta\tilde{v} + \|\tilde{v}\|\Delta\tilde{u}}{\|\tilde{v}\|^2}$

4.2.3 Estabilidad y condición

La *condición* de un problema matemático relaciona su sensibilidad con los cambios en los datos de entrada. Se dice que un cálculo es *numéricamente inestable* si la inexactitud de los valores de entrada se aumenta considerablemente por el método numérico.

Estas ideas pueden estudiarse usando la serie de Taylor de primer orden

$$f(x) = f(\tilde{x}) + f'(\tilde{x})(x - \tilde{x})$$

Esta relación se emplea para estimar el *error relativo* de $f(x)$ como en

$$\frac{f(x) - f(\tilde{x})}{f(x)} \cong \frac{f'(\tilde{x})(x - \tilde{x})}{f(\tilde{x})}$$

El *error relativo* de x está dado por

$$\frac{x - \tilde{x}}{\tilde{x}}$$

Un *número de condición* puede definirse como la razón entre estos errores relativos

$$\text{Número de condición } = \frac{\tilde{x}\, f'(\tilde{x})}{f(\tilde{x})} \tag{4.28}$$

El número de condición proporciona una medida de qué tanto una inexactitud de x se aumenta por $f(x)$. Un valor de 1 nos indica que el error relativo de la función es idéntico al error relativo de x. Un valor mayor que 1 nos señala que el error relativo se amplifica; mientras que para un valor menor que 1 nos dice que se atenúa. En funciones con valores muy grandes se dice que están *mal condicionadas*. Cualquier combinación de los factores en la ecuación (4.28), que aumente el valor numérico del número de condición, tendería a aumentar inexactitudes al calcular $f(x)$.

EJEMPLO 4.7 Número de condición

Planteamiento del problema Calcule e interprete el número de condición para

$$f(x) = \tan x \qquad \text{para } \tilde{x} = \frac{\pi}{2} + 0.1\left(\frac{\pi}{2}\right)$$

$$f(x) = \tan x \qquad \text{para } \tilde{x} = \frac{\pi}{2} + 0.01\left(\frac{\pi}{2}\right)$$

Solución El número de condición se calcula como

$$\text{Número de condición} = \frac{\tilde{x}(1/\cos^2 x)}{\tan \tilde{x}}$$

Para $\tilde{x} = \pi/2 + 0.1(\pi/2)$

$$\text{Número de condición} = \frac{1.7279(40.86)}{-6.314} = -11.2$$

Así, la función está mal condicionada. Para $\tilde{x} = \pi/2 + 0.01\ (\pi/2)$, esta situación es aún peor:

$$\text{Número de condición} = \frac{1.5865(4\ 053)}{-63.66} = -101$$

En este caso, la causa principal del mal condicionamiento parece ser la derivada. Esto tiene sentido, ya que en la vecindad de $\pi/2$, la tangente tiende tanto a infinito positivo como a infinito negativo.

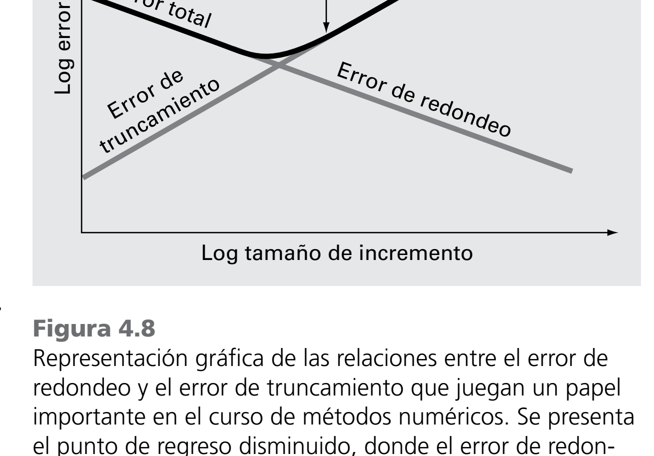


Figura 4.8
Representación gráfica de las relaciones entre el error de redondeo y el error de truncamiento que juegan un papel importante en el curso de métodos numéricos. Se presenta el punto de regreso disminuido, donde el error de redondeo no muestra los beneficios de la reducción del tamaño del incremento.

4.3 ERROR NUMÉRICO TOTAL

El *error numérico total* es la suma de los errores de truncamiento y de redondeo. En general, la única forma para minimizar los errores de redondeo consiste en incrementar el número de cifras significativas en la computadora. Adicionalmente, hemos notado que el error de redondeo *aumentará* debido a la cancelación por resta o debido a que en el análisis aumente el número de cálculos. En contraste, el ejemplo 4.4 demuestra que el error de truncamiento se reduce disminuyendo el tamaño del incremento. Como una disminución al tamaño del incremento puede llevar a una cancelación por resta o a un incremento de los cálculos, los errores de truncamiento *disminuyen* conforme los errores de redondeo se *incrementan*. En consecuencia, se debe afrontar el siguiente dilema: la estrategia para disminuir un componente del error total conduce a un incremento en el otro componente. En un cálculo, se podría disminuir el tamaño del incremento para minimizar los errores de truncamiento únicamente para descubrir que el error de redondeo empieza a dominar la solución y ¡el error total crece! Así, el remedio empieza a ser un problema (figura 4.8). Es un reto determinar el tamaño del incremento apropiado para un cálculo en particular. Se deberá seleccionar un tamaño de incremento grande con la finalidad de disminuir la cantidad de cálculos y errores de redondeo para no tener como consecuencia grandes errores de truncamiento. Si el error total es como se muestra en la figura 4.8, el reto es identificar un punto llamado de regreso disminuido donde los errores de redondeo no muestran los beneficios de la reducción del tamaño del incremento.

En casos reales, sin embargo, tales situaciones son relativamente poco comunes, porque muchas computadoras utilizan suficientes cifras significativas para que los errores de redondeo no predominen. Aunque algunas veces estos errores ocurren y surge una clase de "principio numérico de incertidumbre" que da un límite absoluto sobre la exactitud que puede obtenerse usando ciertos métodos numéricos computarizados. Se explora dicho caso en la siguiente sección.

4.3.1 Análisis de errores de diferenciación numérica

Como se describe en la sección 4.1.3, una aproximación de diferencias centradas de la primera derivada se puede escribir como [ecuación (4.22)]:

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_{i-1})}{2h} - \frac{f^{(3)}(\xi)}{6}h^2 \tag{4.29}$$

Valor verdadero — Aproximación de diferencias finitas — Error de truncamiento

Así, si los dos valores de la función en el numerador de la aproximación de diferencias finitas no tienen error de redondeo, el único error se debe al truncamiento.

Sin embargo, como se están usando computadoras digitales, los valores de las funciones sí incluyen error de redondeo como en

$$f(x_{i-1}) = \tilde{f}(x_{i-1}) + e_{i-1}$$

$$f(x_{i+1}) = \tilde{f}(x_{i+1}) + e_{i+1}$$

donde las $\tilde{f}$ son los valores redondeados de la función, y las e son los errores asociados de redondeo. Sustituyendo estos valores en la ecuación (4.29), se obtiene

$$f'(x_i) = \frac{\tilde{f}(x_{i+1}) - \tilde{f}(x_{i-1})}{2h} + \frac{e_{i+1} - e_{i-1}}{2h} - \frac{f^{(3)}(\xi)}{6}h^2$$

Valor verdadero — Aproximación de diferencias finitas — Error de redondeo — Error de truncamiento

Se puede ver que el error total de la aproximación por diferencias finitas consiste en un error de redondeo que aumenta con el tamaño de los pasos, y un error de truncamiento que disminuye con el tamaño de los pasos.

Suponiendo que el valor absoluto de cada componente del error de redondeo tiene un límite superior ε, el valor absoluto posible de la diferencia $e_{i+1} - e_i$ será 2ε. Suponga además que la tercera derivada tiene un valor absoluto de M. Un límite superior del valor absoluto del error total puede, por lo tanto, representarse como

$$\text{Error total} = \left| f'(x_i) - \frac{\tilde{f}(x_{i+1}) - \tilde{f}(x_{i-1})}{2h} \right| \leq \frac{\varepsilon}{h} + \frac{h^2 M}{6} \tag{4.30}$$

Un tamaño óptimo de paso se puede determinar diferenciando la ecuación (4.30), igualando a cero el resultado y despejando

$$h_{\text{ópt}} = \sqrt[3]{\frac{3\varepsilon}{M}} \tag{4.31}$$

EJEMPLO 4.8 Errores de redondeo y truncamiento en diferenciación numérica

Planteamiento del problema En el ejemplo 4.4 se usó una aproximación de diferencias centradas de $O(h^2)$ para estimar la primera derivada de la siguiente función en $x = 0.5$,

$$f(x) = -0.1x^4 - 0.15x^3 - 0.5x^2 - 0.25x + 1.2$$

Realice el mismo cálculo comenzando con $h = 1$. Luego divida progresivamente el tamaño del incremento entre un factor de 10 para demostrar cómo se vuelve dominante el redondeo al reducir el tamaño del incremento. Relacione sus resultados con la ecuación (4.31). Recuerde que el valor exacto de la derivada es –0.9125.

Solución Se puede desarrollar un programa para realizar los cálculos y graficar los resultados. Para el ejemplo actual, esto se ha hecho con un archivo M de MATLAB. Observe que tanto la función como su derivada analítica se pasan como argumentos. Además, la función genera una gráfica de los resultados.

```
function diffex(func,dfunc,x,n)
format long
dftrue=dfunc(x);
h=1;
H(1)=h;
D(1)=(func(x+h)-func(x-h))/(2*h);
E(1)=abs(dftrue-D(1));
for i=2:n
```

```
    h=h/10;
    H(i)=h;
    D(i)=(func(x+h)-func(x-h))/(2*h);
    E(i)=abs(dftrue-D(i));
  end
  L=[H' D' E']';
  fprintf('  step size   finite difference   true
    error\n');
  fprintf('%14.10f %16.14f %16.13f\n',L);
  loglog(H,E),xlabel('Step Size'),ylabel('Error')
  title('Plot of Error Versus Step Size')
  format short
```

En seguida se puede correr el archivo M usando los siguientes comandos:

```
>> ff=@(x) -0.1*x^4-0.15*x^3-0.5*x^2-0.25*x+1.2;
>> df=@(x) -0.4*x^3-0.45*x^2-x-0.25;
>> diffex(ff,df,0.5,11)
```

Cuando se ejecuta la función, se genera la siguiente salida numérica junto con la gráfica (figura 4.9):

```
 step size   finite difference    true error
1.0000000000 -1.26250000000000 0.3500000000000
0.1000000000 -0.91600000000000 0.0035000000000
0.0100000000 -0.91253500000000 0.0000350000000
0.0010000000 -0.91250035000001 0.0000003500000
0.0001000000 -0.91250000349985 0.0000000034998
0.0000100000 -0.91250000003318 0.0000000000332
0.0000010000 -0.91250000000542 0.0000000000054
0.0000001000 -0.91249999945031 0.0000000005497
0.0000000100 -0.91250000333609 0.0000000033361
0.0000000010 -0.91250001998944 0.0000000199894
0.0000000001 -0.91250007550059 0.0000000755006
```

Los resultados son como se esperaban. Al principio, el redondeo es mínimo, y el estimado está dominado por el error de truncamiento. Por consiguiente, al igual que en la ecuación (4.30), el error total cae por un factor de 100 cada vez que se divide el incremento entre 10. Sin embargo, comenzando en $h = 0.0001$, se ve que el error de redondeo comienza a aumentar y a hacer que disminuya la rapidez de disminución del error. Se llega a un error mínimo en $h = 10^{-6}$. Más allá de este punto, el error aumenta al hacerse dominante el redondeo.

Como se trata de una función fácilmente diferenciable, se puede también investigar si estos resultados son congruentes con la ecuación (4.31). Primero, se puede estimar M evaluando la tercera derivada de la función como

$$M = |f^3(0.5)| = |-2.4(0.5) - 0.9| = 2.1$$

Como MATLAB tiene una precisión de alrededor de 15 a 16 dígitos base 10, una estimación aproximada del límite superior en el redondeo sería alrededor de $\varepsilon = 0.5 \times 10^{-16}$. Sustituyendo estos valores en la ecuación (4.31), se obtiene

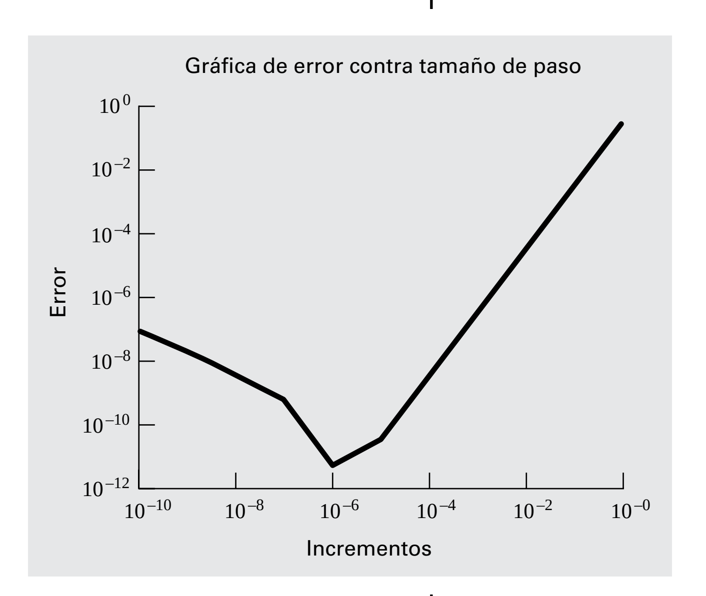




Figura 4.9
Gráfica de error contra tamaño de paso.

$$h_{\text{ópt}} = \sqrt[3]{\frac{3(0.5 \times 10^{-16})}{2.1}}$$

$$= 4.3 \times 10^{-6}$$

que es del mismo orden que el resultado de 1×10^{-6} que se obtuvo con el programa de computación.

4.3.2 Control de errores numéricos

En la mayoría de los casos prácticos, no se conoce el error exacto asociado con el método numérico. Con excepción, claro está, de cuando obtenemos la solución exacta que vuelve innecesaria la aproximación numérica. Por lo tanto, en la mayoría de las aplicaciones en ingeniería debe tenerse algún estimado del error en los cálculos.

No hay una forma sistemática ni general para evaluar el error numérico en todos los problemas. En muchos casos, la estimación del error se basa en la experiencia y en el buen juicio del ingeniero.

Aunque el análisis de error es hasta cierto punto un arte, se sugieren varios lineamientos prácticos de cálculo: lo primero, y principal, implica tratar de evitar la resta de dos números casi iguales. Cuando esto ocurre, casi siempre se pierden cifras significativas. Algunas veces puede reordenarse o reformularse el problema para evitar la cancelación por resta. Y si esto no es posible, se utiliza la aritmética de precisión extendida. Además, cuando se suman o se restan números, es mejor ordenarlos y trabajar primero con los números más pequeños, lo cual evita perder cifras significativas.

Más allá de estas sugerencias de cálculo, se puede intentar predecir el error numérico total usando formulaciones teóricas. La serie de Taylor es la primera herramienta de análisis tanto para el error de truncamiento como para el error de redondeo. Varios ejemplos se han presentado en este capítulo. La predicción del error numérico total es muy complicada para, incluso, un problema de tamaño moderado, y tiende a resultar pesimista. Por lo tanto, únicamente se utiliza para tareas a pequeña escala.

La tendencia es avanzar con los cálculos numéricos e intentar estimar la exactitud de sus resultados. Esto algunas veces se puede hacer observando si los resultados satisfacen alguna condición o ecuación de prueba. O se pueden sustituir los resultados en la ecuación original para verificar si se satisface dicha ecuación.

Por último, usted debería estar preparado para realizar experimentos numéricos que aumenten su conocimiento de los errores de cálculo y de posibles problemas mal condicionados. Tales experimentos pueden consistir en repetir los cálculos con diferentes tamaños de incremento o método, y comparar los resultados. Llega a emplearse un análisis sensitivo para observar cómo la solución cambia cuando se modifican los parámetros del modelo o los valores de entrada. Es factible probar distintos algoritmos numéricos que tengan diferente fundamento matemático, que se basan en distintas estrategias de cálculo o que tengan diferentes características de convergencia y de estabilidad.

Cuando los resultados del cálculo numérico son extremadamente críticos y pueden implicar la pérdida de vidas humanas o tener severas repercusiones económicas, es apropiado tomar precauciones especiales. Esto implicaría el uso de dos o más grupos independientes para resolver el mismo problema y luego comparar los resultados.

El papel de los errores será un tópico de preocupación y análisis en todas las secciones de este libro. Se dejan estas investigaciones en secciones específicas.

4.4 EQUIVOCACIONES, ERRORES DE FORMULACIÓN E INCERTIDUMBRE EN LOS DATOS

Aunque las siguientes fuentes de error no están directamente relacionadas con la mayor parte de los métodos numéricos de este libro, en algunas ocasiones llegan a tener un gran impacto en el éxito al realizar un modelado. Por lo tanto, se deben tener siempre en cuenta cuando se apliquen técnicas numéricas en el contexto de los problemas del mundo real.

4.4.1 Errores por equivocación

A todos nos son familiares los errores por negligencia o por equivocación. En los primeros años de las computadoras, los resultados numéricos erróneos algunas veces se atribuían a las fallas de la propia computadora. En la actualidad esta fuente de error es muy improbable y la mayor parte de las equivocaciones se atribuyen a fallas humanas.

Las equivocaciones llegan a ocurrir a cualquier nivel del proceso de modelación matemática y pueden contribuir con todas las otras componentes del error. Es posible evitarlos únicamente con un sólido conocimiento de los principios fundamentales y mediante el cuidado con el que se enfoque y diseñe la solución del problema.

Las equivocaciones por lo general se pasan por alto en el estudio de un método numérico. Esto se debe sin duda al hecho de que los errores son, hasta cierto punto, inevitables. No obstante, recuerde que hay varias formas con las cuales se puede minimizar su aparición. En particular, los buenos hábitos de programación que se esbozaron en el capítulo 2 son muy útiles para disminuir las equivocaciones. Además, hay formas simples de verificar si un método numérico funciona correctamente. A lo largo del texto, se estudian algunas formas de verificar los resultados de un cálculo numérico.

4.4.2 Errores de formulación

Los *errores de formulación* o *de modelo* pueden atribuirse al sesgo que implica un modelo matemático incompleto. Un ejemplo de un error de formulación insignificante es el hecho de que la segunda ley de Newton no toma en cuenta los efectos relativísticos. Esto no desvirtúa la validez de la solución del ejemplo 1.1, ya que estos errores son mínimos en las escalas de tiempo y espacio asociadas con el problema de la caída del paracaidista.

Sin embargo, suponga que la resistencia del aire no es linealmente proporcional a la velocidad de caída, como en la ecuación (1.7), sino que está en función del cuadrado de la velocidad. Si éste fuera el caso, las soluciones analíticas y numéricas obtenidas en el primer capítulo serían falsas debido al error en la formulación. En algunas aplicaciones de ingeniería del libro se presentan consideraciones adicionales a los errores de formulación. Se debe estar consciente de estos problemas y darse cuenta de que, si se está usando un modelo deficiente, ningún método numérico generará los resultados adecuados.

4.4.3 Incertidumbre en los datos

Algunas veces se introducen errores en un análisis debido a la incertidumbre en los datos físicos obtenidos, sobre los que se basa el modelo. Por ejemplo, suponga que se desea probar el modelo de la caída del paracaidista, haciendo que un individuo salte repetidas veces, midiendo su velocidad después de un intervalo de tiempo específico. Sin duda, se asociaría cada medición con una incertidumbre, ya que el paracaidista caerá con más rapidez en unos saltos que en otros. Estos errores pueden mostrar inexactitud e imprecisión. Si los instrumentos constantemente subevalúan o sobrevalúan las mediciones de la velocidad, se estará tratando con un instrumento inexacto o desviado. Por otro lado, si las medidas son aleatoriamente grandes y pequeñas, entonces se trata de una cuestión de precisión.

Los errores de medición se pueden cuantificar resumiendo los datos con uno o más estadísticos, que den tanta información como sea posible, respecto a características específicas de los datos. Tales estadísticos descriptivos a menudo se seleccionan para obtener 1) la posición del centro de la distribución de los datos y 2) el grado de dispersión de los datos. Como tales, estos estadísticos ofrecen una medida de la desviación e imprecisión, respectivamente. En la parte cinco se regresa el tema de caracterización de incertidumbre de datos.

Aunque se debe estar consciente de los errores por equivocación, de los errores de formulación y de la incertidumbre en los datos, los métodos numéricos utilizados para construir modelos pueden estudiarse, en la mayoría de los casos, en forma independiente de estos errores. Por consiguiente, en la mayor parte de este libro se supondrá que no hay errores por equivocaciones, que el modelo es adecuado y que se está trabajando sin errores en las mediciones de los datos. En estas condiciones es posible estudiar los métodos numéricos sin complicaciones.

PROBLEMAS

4.1 La serie infinita se utiliza para aproximar e^x:

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}$$

a) Muestre que la expansión en serie de Maclaurin es un caso especial de la expansión en la serie de Taylor [ecuación (4.7)] con $x_i = 0$ y $h = x$.

b) Use la serie de Taylor para estimar $f(x) = e^{-x}$ en $x_{i+1} = 1$ para $x_i = 0.2$. Emplee versiones de cero, primero, segundo y tercer orden, y calcule $|\varepsilon_t|$ para cada caso.

4.2 La expansión en serie de Maclaurin para cos x es

$$\cos x = 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \cdots$$

Iniciando con el primer término cos $x = 1$, agregue los términos uno a uno para estimar $\cos(\pi/3)$. Después de que agregue cada uno de los términos, calcule los errores relativos porcentuales verdaderos y aproximados. Use una calculadora para determinar el valor exacto. Agregue términos hasta que el valor absoluto del error aproximado se encuentre dentro de cierto criterio de error, considerando dos cifras significativas.

4.3 Repita los cálculos del problema 4.2, pero ahora usando la expansión de la serie de Maclaurin para sen x, para evaluar el sen$(\pi/3)$.

$$\text{sen } x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$$

4.4 La expansión de la serie de Maclaurin para la arcotangente de x se define para $|x| \leq 1$ como

$$\arctan x = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} x^{2n+1}$$

a) Escriba los primeros cuatro términos ($n = 0\ldots3$).

b) Comenzando con la versión más simple, arctan $x = x$, agregue términos, uno a la vez, para estimar $\arctan(\pi/6)$. Después de agregar cada nuevo término, calcule los errores porcentuales relativos verdaderos y aproximados. Use su calculadora para determinar el valor exacto. Agregue términos hasta que el valor absoluto del estimado de error aproximado caiga por debajo de un criterio de error ajustado a dos cifras significativas.

4.5 Emplee la expansión de la serie de Taylor de cero hasta tercer orden para predecir $f(3)$ si

$$f(x) = 25x^3 - 6x^2 + 7x - 88$$

usando como punto base $x = 1$. Calcule el error relativo porcentual verdadero ε_t para cada aproximación.

4.6 Use la expansión de la serie de Taylor de cero al cuarto orden para estimar $f(2.5)$ si $f(x) = \ln x$ utilizando $x = 1$ como punto base. Calcule el error relativo porcentual verdadero ε_t para cada aproximación. Analice los resultados.

4.7 Utilice aproximaciones en diferencias de $O(h)$ hacia atrás y hacia adelante y una aproximación de diferencia central de $O(h^2)$ para estimar la primera derivada de la función mencionada en el problema 4.5. Evalúe la derivada en $x = 2$ usando un tamaño del incremento $h = 0.2$. Compare los resultados con el valor exacto de las derivadas. Interprete los resultados considerando el término residual de la expansión en la serie de Taylor.

4.8 Con la aproximación en diferencias centrales de $O(h^2)$ estime la segunda derivada de la función examinada en el problema 4.5. Realice la evaluación para $x = 2$ usando un tamaño de incremento $h = 0.25$ y 0.125. Compare lo estimado con el valor exacto de la segunda derivada. Interprete sus resultados considerando el término residual de la expansión en la serie de Taylor.

4.9 La ley de Stefan-Boltzmann se utiliza para estimar la tasa de radiación de energía H desde una superficie, esto es,

$$H = Ae\sigma T^4$$

donde H está en watts, A = área de la superficie (m^2), e = la emisividad que caracteriza la propiedad de emisión de la superficie (adimensional), σ = una constante universal llamada constante de Stefan-Boltzmann (= 5.67×10^{-8} W m^{-2} K^{-4}) y T = temperatura absoluta (K). Determine el error de H para una placa de acero con $A = 0.15$ m^2, $e = 0.90$ y $T = 650 \pm 20$. Compare los resultados con el error exacto. Repita los cálculos pero con $T = 650 \pm 40$. Interprete los resultados.

4.10 Repita el problema 4.9, pero para una esfera de cobre con radio $= 0.15 \pm 0.01$ m, $e = 0.90 \pm 0.05$ y $T = 550 \pm 20$.

4.11 Recuerde que la velocidad de caída del paracaidista puede calcularse con [ecuación (1.10)]

$$v(t) = \frac{gm}{c}(1 - e^{-(c/m)t})$$

Use un análisis de error de primer orden para estimar el error de v para $t = 6$, si $g = 9.81$ y $m = 50$, pero $c = 12.5 \pm 1.5$.

4.12 Repita el problema 4.11 con $g = 9.8$, $t = 6$, $c = 12.5 \pm 1.5$ y $m = 50 \pm 2$.

4.13 Evalúe e interprete los números de condición para

a) $f(x) = \sqrt{|x-1|} + 1$ para $x = 1.00001$

b) $f(x) = e^{-x}$ para $x = 10$

c) $f(x) = \sqrt{x^2+1} - x$ para $x = 300$

d) $f(x) = \dfrac{e^x - 1}{x}$ para $x = 0.001$

e) $f(x) = \dfrac{\text{sen } x}{1 + \cos x}$ para $x = 1.0001\pi$

4.14 Empleando las ideas de la sección 4.2, muestre las relaciones de la tabla 4.3.

4.15 Muestre que la ecuación (4.4) es exacta para todos los valores de x, si $f(x) = ax^2 + bx + c$.

4.16 La fórmula de Manning para un canal rectangular se escribe como

$$Q = \frac{1}{n}\frac{(BH)^{5/3}}{(B+2H)^{2/3}}\sqrt{S}$$

donde Q = flujo (m^3/s), n = coeficiente de rugosidad, B = ancho (m), H = profundidad (m) y S = pendiente. Aplique la fórmula para un arroyo donde se conoce que el ancho = 20 m y la profundidad = 0.3 m. Por desgracia conocemos el coeficiente de rugosidad y la pendiente con una precisión de sólo ±10%. Es decir, la rugosidad tiene un valor de 0.03 con un rango de 0.027 a 0.033, y la pendiente es 0.0003 con un rango de 0.00027 a 0.00033. Use un análisis de error de primer orden para determinar la sensibilidad en la predicción del flujo para cada uno de esos dos factores. ¿Cuál variable se debería intentar medir para una mejor precisión?

4.17 Si $|x| < 1$, se sabe que

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots$$

Repita el problema 4.1 para esta serie con $x = 0.1$.

4.18 Un misil sale de la Tierra con una velocidad inicial v_0 formando con la vertical un ángulo ϕ_0 como se muestra en la figura P4.18. La altitud máxima deseada es αR, donde R es el radio de la Tierra. Usando las leyes de la mecánica se demuestra que

$$\text{sen } \phi_0 = (1+\alpha)\sqrt{1 - \frac{\alpha}{1+\alpha}\left(\frac{v_e}{v_0}\right)^2}$$

donde v_e es la velocidad de escape del misil. Se quiere disparar el misil y alcanzar la velocidad máxima proyectada con una exactitud de ±2%. Determine el rango de valores de ϕ_0 si $v_e/v_0 = 2$ y $\alpha = 0.25$.

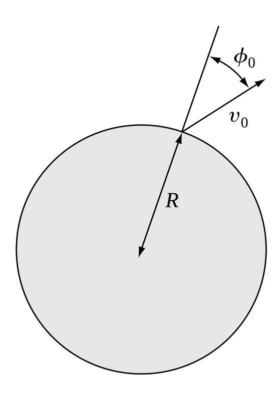


Figura P4.18

4.19 Para calcular las coordenadas espaciales de un planeta tenemos que resolver la función

$$f(x) = x - 1 - 0.5 \text{ sen } x$$

Sea $a = x_i = \pi/2$ en el intervalo $[0, \pi]$ el punto base. Determine la expansión de la serie de Taylor de orden superior que da un error máximo de 0.015 en el intervalo dado. El error es igual al valor absoluto de la diferencia entre la función dada y la expansión de la serie de Taylor especificada. (*Sugerencia:* Resuelva gráficamente.)

4.20 Considere la función $f(x) = x^3 - 2x + 4$ en el intervalo $[-2, 2]$ con $h = 0.25$. Use las aproximaciones en diferencias finitas hacia adelante, hacia atrás y centrada para la primera y segunda derivadas, e ilustre gráficamente qué aproximación es más exacta. Grafique las tres aproximaciones a la primera derivada por diferencias finitas, junto con los valores exactos, y haga lo mismo con la segunda derivada.

4.21 Derive la ecuación (4.31).

4.22 Repita el ejemplo 4.8 pero para $f(x) = \cos(x)$ en $x = \pi/6$.

4.23 Repita el ejemplo 4.8, pero con diferencias divididas hacia adelante [ecuación (4.17)].

4.24 Desarrolle un programa bien estructurado para calcular la expansión de la serie de Maclaurin para la función coseno como se describe en el problema 4.2. La función debe tener las siguientes características:

- Itere hasta que el error relativo caiga por debajo de un criterio de detención (`es`) o exceda un número máximo de iteraciones (`maxit`). Permita que el usuario especifique valores para estos parámetros.
- Incluya valores por defecto de `es` (= 0.000001) y `maxit` (= 100) en caso de que no sean especificados por el usuario.
- Regrese el estimado de $\cos(x)$, el valor relativo aproximado, el número de iteraciones y el valor relativo verdadero (que usted puede calcular con base en la función coseno interconstruida).

EPÍLOGO: PARTE UNO

PT1.4 ALTERNATIVAS

Los métodos numéricos son científicos en el sentido de que representan técnicas sistemáticas para resolver problemas matemáticos. Sin embargo, hay cierto grado de arte, juicios subjetivos y *conveniencias*, relacionadas con su uso efectivo en la ingeniería práctica. Para cada problema, se enfrenta uno con varios métodos numéricos alternativos y con muchos tipos diferentes de computadoras. Así, la elegancia y la eficiencia de las diferentes maneras de abordar los problemas varían de una persona a otra y se correlacionan con la habilidad de hacer una elección prudente. Por desgracia, como sucede con cualquier proceso intuitivo, los factores que influyen en dicha elección son difíciles de comunicar. Estas habilidades pueden descubrirse y desarrollarse sólo mediante la experiencia. Como tales habilidades desempeñan un papel muy importante en el uso efectivo de los métodos, se presenta esta sección como una introducción a algunas de las *alternativas* que se deben considerar cuando se seleccione un método numérico y las herramientas para su realización. Se espera que el siguiente análisis influencie su orientación cuando estudie el material subsecuente. También, que usted consulte nuevamente el material cuando enfrente distintas alternativas en el resto del libro.

1. *Tipo de problema matemático.* Como se definió previamente en la figura PT.1.2, en este libro se analizan varios tipos de problemas matemáticos.

a) Raíces de ecuaciones.
b) Sistemas de ecuaciones algebraicas lineales simultáneas.
c) Optimización.
d) Ajuste de curvas.
e) Integración numérica.
f) Ecuaciones diferenciales ordinarias.
g) Ecuaciones diferenciales parciales.

Probablemente el lector se encontrará con algunos aspectos básicos sobre la aplicación de los métodos numéricos al enfrentarse con problemas específicos en algunas de esas áreas. Los métodos numéricos son necesarios, ya que los problemas planteados no se pueden resolver en su totalidad usando técnicas analíticas. Deberá estar consciente de que en las actividades profesionales se encontrarán problemas en las áreas ya mencionadas. Por lo que el estudio de los métodos numéricos y la selección de un equipo de cómputo deben, al menos, considerar esos tipos de problemas básicos. Problemas más avanzados quizá requieran de capacidades en otras áreas como la aproximación funcional, las ecuaciones integrales, etc. Estas áreas requieren de una gran potencia computacional o de métodos avanzados que no se cubren en este texto. Se recomienda consultar algunas referencias como Carnahan, Luther y Wilkes (1969); Hamming (1973); Ralston y Rabinowitz (1978); Burden y Faires (2005), y Moler (2004) para problemas que van más allá del contenido de este libro. Además, al final de cada parte de este texto se ofrece un resumen y las referencias para los métodos numéricos avanzados con la finalidad de encauzar al lector en el estudio de este tipo de métodos numéricos.

2. *Tipo, disponibilidad, precisión, costo y velocidad de una computadora.* Se puede tener la oportunidad de trabajar con varias herramientas de cómputo, que van desde una calculadora de bolsillo hasta una supercomputadora. Cualquiera de estas herramientas se puede usar para implementar un método numérico (incluyendo simple papel y lápiz). En general, no se trata de extremar la capacidad, sino más bien evaluar costo, conveniencia, velocidad, seguridad, exactitud y precisión. Aunque cada una de las herramientas seguirán teniendo utilidad, los grandes avances recientes en el funcionamiento de las computadoras personales han tenido un gran impacto en la profesión del ingeniero. Se espera que esta revolución siga extendiéndose conforme continúen los avances tecnológicos, ya que las computadoras personales ofrecen una excelente combinación de conveniencia, costo, precisión, velocidad y capacidad de almacenamiento. Más aún, se pueden usar fácilmente en la mayoría de los problemas prácticos de ingeniería.

3. *Costo de desarrollo de programas contra costo de software contra costo de tiempo de ejecución*. Una vez que los tipos de problemas matemáticos que deberán resolverse se hayan identificado y el sistema de cómputo se haya seleccionado, se considerarán los costos del software y del tiempo de ejecución. El desarrollo de software llega a representar un trabajo adicional en muchos proyectos de ingeniería y, por lo tanto, tener un costo sustancial. A este respecto, es importante que conozca bien los aspectos teóricos y prácticos de los métodos numéricos relevantes. Además, debe familiarizarse con el desarrollo del software profesional. Existe software de bajo costo disponible para implementar métodos numéricos, el cual es fácilmente adaptado a una amplia variedad de problemas.
4. *Características de los métodos numéricos*. Si el costo de una computadora y de sus programas es alto, o si la disponibilidad de la computadora es limitada (por ejemplo, en sistemas de tiempo compartido), la manera de escoger cuidadosamente el método numérico ayudará a adaptarse a tal situación. Por otro lado, si el problema aún se encuentra en una etapa experimental, donde el acceso y el costo de una computadora no presenta problemas, entonces es posible seleccionar un método numérico que siempre trabaje, aunque quizá no sea, computacionalmente hablando, el más eficiente. Los métodos numéricos disponibles para resolver un tipo particular de problema implican todos los factores mencionados, además de:

   a) *Número de condiciones iniciales o de puntos de partida*. Algunos de los métodos numéricos para encontrar raíces de ecuaciones, o para la solución de ecuaciones diferenciales, requieren que el usuario especifique las condiciones iniciales o puntos de partida. Los métodos simples requieren en general de un valor, mientras que los métodos complicados tal vez requieran más de un valor. Las ventajas de los métodos complicados, que son computacionalmente eficientes, llegan a compensar requerimientos de puntos de partida múltiples. Debe echar mano de su experiencia y buen juicio para estimar las alternativas que tomará en cada problema en particular.
   b) *Velocidad de la convergencia.* Ciertos métodos numéricos convergen más rápido que otros. No obstante, la convergencia rápida puede requerir de puntos iniciales más adecuados y de programación más compleja, que un método donde la convergencia es lenta. De nueva cuenta deberá usar su propio criterio y la experiencia para seleccionar el método. ¡Lo más rápido no siempre es lo mejor!
   c) *Estabilidad*. Algunos métodos numéricos usados para encontrar raíces de ecuaciones o para resolver sistemas de ecuaciones lineales llegan a diverger en vez de converger a la respuesta correcta. ¿Por qué existe esta posibilidad al enfrentarse con problemas de diseño o de planeación? La respuesta es que tales métodos pueden ser altamente eficientes para determinados problemas; por lo tanto, surgen de nuevo las alternativas. Deben decidir si las condiciones del problema justifican el empleo de un método que quizá no siempre converge.
   d) *Exactitud y precisión*. Algunos de los métodos numéricos son más exactos y precisos que otros. Como ejemplo se tienen las diferentes ecuaciones usadas en la integración numérica. En general, es posible mejorar el funcionamiento de un método de poca exactitud disminuyendo el tamaño del incremento o aumentando el número de aplicaciones en un intervalo dado. ¿Resultará mejor usar un método poco exacto con un tamaño de incremento pequeño o un método de gran exactitud con un tamaño de incremento grande? La pregunta se debe analizar en cada caso específico, tomando en cuenta factores adicionales como el costo y la facilidad de programación. Además, se deben tomar en consideración los errores de redondeo cuando se utilizan métodos de baja exactitud en forma repetida, y cuando la cantidad de cálculos es grande. Aquí, el número de cifras significativas empleadas por la computadora llega a ser el factor decisivo.
   e) *Gama de aplicaciones*. Algunos métodos numéricos se aplican sólo a ciertas clases de problemas o a problemas que satisfacen ciertas restricciones matemáticas. Otros métodos no se ven afectados por estas restricciones. Entonces, deberá evaluar si vale la pena desarrollar programas que emplean técnicas apropiadas únicamente para un número limitado de problemas. El hecho de que tales técnicas sean ampliamente usadas indica que tienen ventajas que a menudo superan a las desventajas. De hecho es necesario evaluar las alternativas.

f) *Requisitos especiales*. Algunas técnicas numéricas tratan de incrementar la exactitud y la velocidad de convergencia usando información especial o adicional. Un ejemplo sería el uso de valores estimados o teóricos de errores que permiten mejorar la exactitud. Sin embargo, estas mejorías, en general, no se logran sin algunos inconvenientes, como mayores costos computacionales o el incremento en la complejidad del programa.

g) *Esfuerzo de programación necesario*. Los esfuerzos para mejorar la velocidad de convergencia, estabilidad y exactitud pueden ser creativos e ingeniosos. Cuando se realizan mejoras sin aumentar la complejidad de la programación, entonces se considera que estas mejoras son excelentes y quizá encuentren un uso inmediato en la ingeniería. No obstante, si éstas requieren de programas más complejos, se enfrentarían a situaciones alternativas que pueden favorecer o no el nuevo método.

Resulta claro que el análisis anterior relacionado con la elección de un método numérico se reduce sólo a costo y exactitud. Los costos son los del tiempo de cómputo y el desarrollo de programas. La exactitud apropiada es una cuestión de ética y de juicio profesional.

5. *Comportamiento matemático de la función, la ecuación o los datos.* Al seleccionar un método numérico en particular, un tipo de computadora y un tipo de programas, se debe tomar en cuenta la complejidad de las funciones, las ecuaciones o los datos. Las ecuaciones simples y los datos uniformes se tratan apropiadamente mediante algoritmos numéricos simples y con computadoras de bajo costo. Sucede lo contrario con las ecuaciones complicadas y los datos que presentan discontinuidades.
6. *Facilidad de aplicación (¿amigable para el usuario?).* Algunos métodos numéricos son fáciles de aplicar; otros son difíciles. Esto es una consideración cuando se tenga que elegir un método sobre otro. La misma idea se aplica a las decisiones que tienen que ver con los costos de desarrollar un programa *versus* el software desarrollado profesionalmente. Podría requerirse un esfuerzo considerable para convertir un programa difícil en otro que sea amigable para el usuario. En el capítulo 2 se introdujeron formas de hacer esto, y se emplean a lo largo del libro.
7. *Mantenimiento.* Los programas para resolver problemas de ingeniería requieren de mantenimiento, porque durante las aplicaciones ocurren, en forma invariable, dificultades. El mantenimiento puede requerir un cambio en el código del programa o la expansión de la documentación. Los programas y los algoritmos numéricos simples son más fáciles de mantener.

Los siguientes capítulos muestran el desarrollo de varios tipos de métodos numéricos para una variedad de problemas matemáticos. Se ofrecen, en cada capítulo, varios métodos alternativos. Se presentan estos métodos (en vez de un método escogido por los autores), ya que no existe uno que sea "el mejor" de todos. No hay métodos "mejores", existen alternativas con ventajas y desventajas que se deben tomar en consideración cuando se aplica un método a un problema práctico. Al final de cada parte del libro, se encontrará una tabla que destaca las ventajas y desventajas inherentes a cada método. Esta tabla debe ayudarle a seleccionar el procedimiento numérico adecuado para el contexto de su problema específico.

PT1.5 RELACIONES Y FÓRMULAS IMPORTANTES

En la tabla PT1.2 se resume información importante que se presentó en la parte uno. La tabla es útil para tener un acceso rápido a las relaciones y fórmulas más importantes. El epílogo de cada parte del libro contiene un resumen como éste.

PT1.6 MÉTODOS AVANZADOS Y REFERENCIAS ADICIONALES

El epílogo de cada parte del libro también incluye una sección diseñada para facilitar y fomentar el estudio de métodos numéricos adicionales. Dicha sección proporciona algunas referencias de otros libros sobre el tema, así como de material relacionado con métodos más avanzados.[1]

[1] Aquí, los libros se *referencian* sólo por autor. Al final del texto se incluye una bibliografía completa.

Tabla PT1.2 Resumen de información importante presentada en la parte uno.

Definiciones de error	
Error verdadero	E_t = valor verdadero – valor aproximado
Error relativo porcentual verdadero	$\varepsilon_t = \dfrac{\text{valor verdadero} - \text{valor aproximado}}{\text{valor verdadero}} 100\%$
Error relativo porcentual aproximado	$\varepsilon_a = \dfrac{\text{aproximación presente} - \text{aproximación anterior}}{\text{aproximación presente}} 100\%$
Criterio de paro	Terminar los cálculos cuando $\varepsilon_a < \varepsilon_s$ donde ε_s es el error relativo porcentual deseado
Serie de Taylor	
Expansión de la serie de Taylor	$f(x_{i+1}) = f(x_i) + f'(x_i)h + \dfrac{f''(x_i)}{2!}h^2 + \dfrac{f'''(x_i)}{3!}h^3 + \cdots + \dfrac{f^{(n)}(x_i)}{n!}h^n + R_n$ donde
Residuo	$R_n = \dfrac{f^{(n+1)}(\xi)}{(n+1)!}h^{n+1}$ o $R_n = O(h^{n+1})$
Diferenciación numérica	
Primera diferencia finita dividida hacia delante	$f'(x_i) = \dfrac{f(x_{i+1}) - f(x_i)}{h} + O(h)$ (Otras diferencias divididas se resumen en los capítulos 4 y 23.)
Propagación del error	
Para n variables independientes $x_1, x_2, \ldots, x_n$ con errores $\Delta\tilde{x}_1, \Delta\tilde{x}_2, \ldots \Delta\tilde{x}_n$, el error en la función f se estima mediante	$\Delta f = \left\lvert\dfrac{\partial f}{\partial x_1}\right\rvert \Delta\tilde{x}_1 + \left\lvert\dfrac{\partial f}{\partial x_2}\right\rvert \Delta\tilde{x}_2 + \cdots + \left\lvert\dfrac{\partial f}{\partial x_n}\right\rvert \Delta\tilde{x}_n$

Para ampliar los antecedentes mencionados en la parte uno, existen diversos manuales sobre programación. Sería difícil mencionar todos los excelentes libros y manuales que corresponden a lenguajes y computadoras específicos. Además quizá ya se tenga material sobre estudios previos de la programación. No obstante, si ésta es su primera experiencia con computadoras, el profesor o sus compañeros de semestre avanzados le darían al usuario recomendaciones acerca de las bibliografías para las máquinas y los lenguajes disponibles en su escuela.

Para el análisis de errores, cualquier buen libro de introducción al cálculo incluirá material complementario relacionado, tal como las series de Taylor. Las obras de Swokowski (1979), Thomas y Finney (1979), y Simmons (1985) ofrecen una teoría comprensible de estos temas. Taylor (1982), además, presenta una excelente introducción al análisis del error.

Finalmente, aunque esperamos que nuestro libro le sea útil, es siempre bueno consultar otras fuentes cuando trate de dominar un nuevo tema. Burden y Faires (2005); Ralston y Rabinowitz (1978); Hoffman (1992), y Carnahan, Luther y Wilkes (1969) proporcionan explicaciones amplias de la mayoría de métodos numéricos. Otros excelentes libros sobre este tema son Gerald y Wheatley (2004), y Cheney y Kincaid (2008). Además, Press *et al.* (2007) incluyen algoritmos para implementar diversos métodos, y Moler (2004) y Chapra (2007) versan sobre métodos numéricos con MATLAB.

PARTE DOS



RAÍCES DE ECUACIONES

PT2.1 MOTIVACIÓN

Desde hace años usted aprendió a usar la fórmula cuadrática:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{PT2.1}$$

para resolver

$$f(x) = ax^2 + bx + c = 0 \tag{PT2.2}$$

A los valores calculados con la ecuación (PT2.1) se les llama las "raíces" de la ecuación (PT2.2), que representan los valores de x que hacen a la ecuación (PT2.2) igual a cero. Por lo tanto, se define la raíz de una ecuación como el valor de x que hace $f(x) = 0$. Debido a esto, algunas veces a las raíces se les conoce como *ceros* de la ecuación.

Aunque la fórmula cuadrática es útil para resolver la ecuación (PT2.2), existen muchas funciones donde las raíces no se pueden determinar tan fácilmente. En estos casos, los métodos numéricos descritos en los capítulos 5, 6 y 7 proporcionan medios eficientes para obtener la respuesta.

PT2.1.1 Métodos para la determinación de raíces sin emplear computadoras

Antes de la llegada de las computadoras digitales se disponía de una serie de métodos para encontrar las raíces de ecuaciones algebraicas y trascendentes. En algunos casos, las raíces se obtenían con métodos directos, como se hace con la ecuación (PT2.1). Sin embargo existen ecuaciones como ésta que se resuelven directamente y aparecen muchas más en las que no es posible encontrar su solución. Por ejemplo, incluso una función tan simple como $f(x) = e^{-x} - x$ no se puede resolver en forma analítica. En tales casos, la única alternativa es una técnica con solución aproximada.

Un método para obtener una solución aproximada consiste en graficar la función y determinar dónde cruza el eje de las x. Este punto, que representa el valor de x para el cual $f(x) = 0$, es la raíz. Las técnicas gráficas se exponen al principio de los capítulos 5 y 6.

Aunque los métodos gráficos son útiles en la obtención de estimaciones de las raíces, tienen el inconveniente de que son poco precisos. Un método alternativo es el de prueba y error. Esta "técnica" consiste en elegir un valor de x y evaluar si $f(x)$ es cero. Si no es así (como sucederá en la mayoría de los casos) se hace otra elección y se evalúa nuevamente $f(x)$ para determinar si el nuevo valor ofrece una mejor aproximación de la raíz. El proceso se repite hasta que se obtenga un valor que proporcione una $f(x)$ cercana a cero.

Estos métodos fortuitos, evidentemente, son ineficientes e inadecuados para las exigencias de la ingeniería. Las técnicas descritas en la parte dos representan alternativas que no sólo aproximan sino que emplean estrategias sistemáticas para dirigirse a la raíz verdadera. Tal como se presenta en las páginas siguientes, la combinación de estos métodos sistemáticos con la computadora hacen que la solución de la mayoría de los problemas de raíces de ecuaciones sea una tarea sencilla y eficiente.

PT2.1.2 Raíces de ecuaciones y la práctica en ingeniería

Aunque las raíces de ecuaciones aparecen en el contexto de diversos problemas, son frecuentes en el área de diseño en ingeniería. En la tabla PT2.1 se muestra un conjunto de principios fundamentales que se utilizan comúnmente en trabajos de diseño. Como se expuso en el capítulo 1, las ecuaciones matemáticas o modelos provenientes de estos principios se utilizan para predecir los valores de

variables dependientes en función de variables independientes y los valores de parámetros. Observe que en cada caso las variables dependientes representan el estado o desempeño del sistema; mientras que los parámetros representan sus propiedades o su composición.

Un ejemplo de tales modelos es la ecuación obtenida a partir de la segunda ley de Newton, usada en el capítulo 1 para la velocidad del paracaidista:

$$v = \frac{gm}{c}\left(1 - e^{-(c/m)t}\right) \tag{PT2.3}$$

donde la velocidad v = la variable dependiente, el tiempo t = la variable independiente, la constante de gravitación g = una función de fuerza y el coeficiente de resistencia c y la masa m son los parámetros. Si se conocen los parámetros, la ecuación (PT2.3) se utiliza para predecir la velocidad del paracaidista como una función del tiempo. Estos cálculos se pueden llevar a cabo de manera directa, ya que v se expresa *explícitamente* como una función del tiempo. Es decir, queda despejada en el lado izquierdo del signo igual.

No obstante, suponga que se tiene que determinar el coeficiente de resistencia de un paracaidista con una masa dada, para alcanzar una velocidad determinada en un periodo preestablecido. Aunque la ecuación (PT2.3) ofrece una representación matemática de la interrelación entre las variables del modelo y los parámetros, no es posible obtener explícitamente el coeficiente de resistencia. Trate de hacerlo. No hay forma de reordenar la ecuación para despejar el parámetro c. En tales casos, se dice que c está en forma *implícita*.

Esto representa un verdadero dilema, ya que en muchos de los problemas de diseño en ingeniería hay que especificar las propiedades o la composición de un sistema (representado por sus parámetros) para asegurar que esté funcionando de la manera deseada (representado por las variables). Así, a menudo dichos problemas requieren la determinación de parámetros implícitos.

La solución del dilema es proporcionada por los métodos numéricos para raíces de ecuaciones. Para resolver el problema con métodos numéricos es conveniente reexpresar la ecuación (PT2.3), esto se logra restando la variable dependiente v de ambos lados de la ecuación,

$$f(c) = \frac{gm}{c}\left(1 - e^{-(c/m)t}\right) - v \tag{PT2.4}$$

Por lo tanto, el valor de c que hace $f(c) = 0$ es la raíz de la ecuación. Este valor también representa el coeficiente de arrastre que resuelve el problema de diseño.

En la parte dos de este libro se analiza una gran variedad de métodos numéricos y gráficos para determinar raíces de relaciones como en la ecuación (PT2.4). Dichas técnicas se pueden apli-

Tabla PT2.1 Principios fundamentales usados en los problemas de ingeniería.

Principio fundamental	Variable dependiente	Variable independiente	Parámetros
Balance de calor	Temperatura	Tiempo y posición	Propiedades térmicas del material y geometría del sistema
Balance de masa	Concentración o cantidad de masa	Tiempo y posición	El comportamiento químico del material: coeficientes de transferencia de masa y geometría del sistema
Balance de fuerzas	Magnitud y dirección de fuerzas	Tiempo y posición	Resistencia del material, propiedades estructurales y geometría del sistema
Balance de energía	Cambios en los estados de energía cinética y potencial de un sistema	Tiempo y posición	Propiedades térmicas, masa del material y geometría del sistema
Leyes de Newton del movimiento	Aceleración, velocidad y posición	Tiempo y posición	Masa del material, geometría del sistema y parámetros disipadores, como fricción y resistencia
Leyes de Kirchhoff	Corriente y voltaje en circuitos eléctricos	Tiempo	Propiedades eléctricas del sistema, como resistencia, capacitancia e inductancia

car a problemas de diseño en ingeniería con base en los principios fundamentales dados en la tabla PT2.1, así como a muchos problemas que se encuentran de manera rutinaria en la práctica de la ingeniería.

PT2.2 ANTECEDENTES MATEMÁTICOS

En la mayoría de las áreas mencionadas en este libro existen algunos prerrequisitos matemáticos necesarios para dominar el tema. Por ejemplo, los conceptos de estimación del error y expansión de la serie de Taylor, analizados en los capítulos 3 y 4, tienen relevancia directa en nuestro estudio de las raíces de ecuaciones. Además, anteriormente ya se mencionaron los términos: ecuaciones “algebraicas” y “trascendentes”. Resulta útil definir formalmente dichos términos y estudiar cómo se relacionan en esta parte del libro.

Por definición, una función dada por $y = f(x)$ es algebraica si se expresa de la forma:

$$f_n y^n + f_{n-1} y^{n-1} + \cdots + f_1 y + f_0 = 0 \tag{PT2.5}$$

donde f_i es un polinomio de i-ésimo orden en x. Los *polinomios* son un tipo de funciones algebraicas que generalmente se representan como:

$$f_n(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \tag{PT2.6}$$

donde n es el *grado* del polinomio y las a son constantes. Algunos ejemplos específicos son:

$$f_2(x) = 1 - 2.37x + 7.5x^2 \tag{PT2.7}$$

y

$$f_6(x) = 5x^2 - x^3 + 7x^6 \tag{PT2.8}$$

Las funciones *trascendentes* son funciones que no son algebraicas. Comprenden las funciones trigonométricas, las funciones exponenciales, las funciones logarítmicas y otras menos familiares. Algunos ejemplos son:

$$f(x) = \ln x^2 - 1 \tag{PT2.9}$$

y

$$f(x) = e^{-0.2x} \text{ sen } (3x - 0.5) \tag{PT2.10}$$

Las raíces de las ecuaciones pueden ser reales o complejas. Aunque hay algunos casos en que las raíces complejas de funciones no polinomiales son de interés, esta situación es menos común que en polinomios. En consecuencia, los métodos numéricos estándares para encontrar raíces se encuentran en dos áreas de problemas relacionados, pero fundamentalmente distintos:

1. *La determinación de raíces reales de ecuaciones algebraicas y trascendentes.* Dichas técnicas se diseñaron para determinar el valor de una sola raíz real basándose en un conocimiento previo de su posición aproximada.
2. *La determinación de todas las raíces reales y complejas de polinomios.* Estos métodos están diseñados especialmente para polinomios; determinan sistemáticamente todas las raíces del polinomio en lugar de sólo una raíz real dada una posición aproximada.

En este libro se estudian ambas, los capítulos 5 y 6 se dedican a la primera área y el capítulo 7 se ocupa de los polinomios.

PT2.3 ORIENTACIÓN

Antes de proceder con los métodos numéricos para determinar raíces de ecuaciones, será útil dar alguna orientación. El siguiente material intenta dar una visión general de los temas de la parte dos. Además, se han incluido algunos objetivos que orientarán al lector en su estudio del material.

PT2.3.1 Alcance y presentación preliminar

La figura PT2.1 es una representación esquemática de la organización de la parte dos. Examine esta figura cuidadosamente, iniciando en la parte de arriba y avanzando en el sentido de las manecillas del reloj.

Después de la presente introducción, el *capítulo 5* se dedica a los *métodos cerrados*, *que usan intervalos*, para encontrar raíces. Estos métodos empiezan con intervalos que encierran o contienen a la raíz, y después reducen sistemáticamente el tamaño del intervalo. Se estudian dos métodos específicos: el de *bisección* y el de la *falsa posición*. Los métodos gráficos sirven para dar una comprensión visual de las técnicas. Se desarrollan formulaciones del error para ayudar a determinar el trabajo computacional que se requiere para estimar la raíz con un nivel de precisión especificado previamente.

En el *capítulo 6* se tratan los *métodos abiertos*. Estos métodos también emplean iteraciones sistemáticas de prueba y error; pero no requieren que el intervalo inicial encierre a la raíz. Se descubrirá que estos métodos, en general, son más eficientes computacionalmente que los métodos cerrados, aunque no siempre funcionan. Se analizan los métodos de *iteración de un punto fijo*, de *Newton-Raphson* y de la *secante*. Los métodos gráficos sirven para dar una idea geométrica en los casos donde los métodos abiertos no funcionan. Se desarrollan las fórmulas que proporcionan una idea de qué tan rápido los métodos abiertos *convergen* a la raíz. Se describe un enfoque avanzado, el *método de Brent*, que combina la confiabilidad del uso de métodos cerrados con la rapidez de los métodos abiertos. Además, se explica la forma de extender el método de Newton-Raphson *para sistemas de ecuaciones no lineales*.

El *capítulo 7* está dedicado a encontrar las *raíces de polinomios*. Después de las secciones anteriores sobre polinomios, se estudian los métodos convencionales (en particular los métodos abiertos del capítulo 6). Se describen dos métodos especiales para localizar raíces de polinomios: los métodos de Müller y Bairstow. El capítulo termina con información acerca de la búsqueda de raíces con Excel, MATLAB y Mathcad.

En el *capítulo 8* se extienden los conceptos anteriores a los problemas reales de ingeniería. Se emplean aplicaciones a la ingeniería para ilustrar las ventajas y desventajas de cada uno de los métodos, proporcionando una visión de cómo se aplican las técnicas en la práctica profesional. Las aplicaciones también destacan las alternativas (estudiadas en la parte uno) asociadas con cada uno de los métodos.

Se incluye un epílogo al final de la parte dos. Éste contiene una detallada comparación de los métodos analizados en los capítulos 5, 6 y 7. Esta comparación comprende una descripción de las alternativas relacionadas con el uso apropiado de cada técnica. Esta sección proporciona también un resumen de las fórmulas importantes, junto con referencias para algunos de los métodos que van más allá del alcance de este texto.

PT2.3.2 Metas y objetivos

Objetivos de estudio Después de terminar la parte dos usted debe tener la suficiente información para abordar con éxito una amplia variedad de problemas de ingeniería, relacionados con las raíces de ecuaciones. En general, ya dominará las técnicas, ya habrá aprendido a determinar su confiabilidad y tendrá la capacidad de elegir el mejor método (o métodos) para cualquier problema particular. Además de estas metas generales, deberá haber asimilado los conceptos específicos de la tabla PT2.2 para comprender mejor el material de la parte dos.

Objetivos de cómputo El libro proporciona software y algoritmos sencillos para implementar las técnicas analizadas en la parte dos. Todos tienen utilidad como herramientas del aprendizaje.

Se presentan directamente pseudocódigos para varios métodos en el texto. Esta información le permitirá ampliar su biblioteca de software para contar con programas que son más eficientes que el método de bisección. Por ejemplo, tal vez usted desee tener sus propios programas para las técnicas de la falsa posición, de Newton-Raphson y de secante, las cuales a menudo son más eficientes que el método de bisección.

Finalmente, los paquetes de software como Excel, MATLAB y Mathcad tienen poderosas capacidades para localizar raíces. Puede usar esta parte del libro para empezar a familiarizarse con estas capacidades.

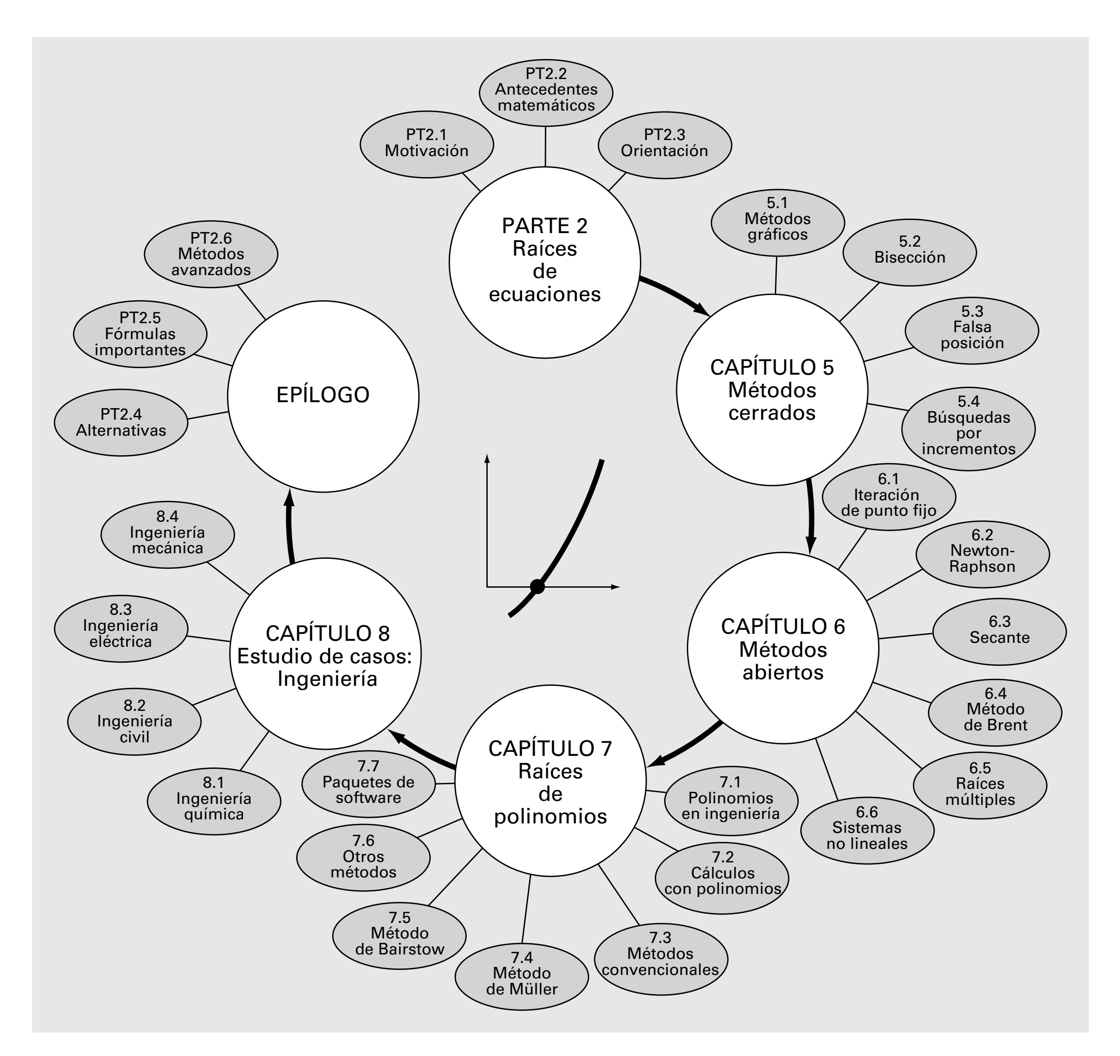


Figura PT2.1
Esquema de la organización del material de la parte dos: Raíces de ecuaciones.

Tabla PT2.2 Objetivos específicos de estudio de la parte dos.

1. Comprender la interpretación gráfica de una raíz
2. Conocer la interpretación gráfica del método de falsa posición y por qué, en general, es mejor que el método de bisección
3. Entender la diferencia entre los métodos cerrados y los métodos abiertos para la localización de las raíces
4. Entender los conceptos de convergencia y de divergencia; usar el método gráfico de las dos curvas para tener una idea visual de los conceptos
5. Saber por qué los métodos cerrados siempre convergen, mientras que los métodos abiertos algunas veces pueden diverger
6. Observar que la convergencia en los métodos abiertos es más segura si el valor inicial está cercano a la raíz verdadera
7. Entender los conceptos de convergencia lineal y cuadrática, así como sus implicaciones en la eficiencia de los métodos de iteración de punto fijo y de Newton-Raphson
8. Conocer las diferencias fundamentales entre el método de la falsa posición y el método de la secante, y cómo se relacionan con la convergencia
9. Entender cómo el método de Brent combina la confiabilidad de la bisección con la rapidez de los métodos abiertos
10. Comprender los problemas que presentan raíces múltiples y las modificaciones que se pueden hacer para reducir dichos problemas
11. Saber cómo extender el método de Newton-Raphson de una sola ecuación no lineal con el propósito de resolver sistemas de ecuaciones no lineales

CAPÍTULO 5

Métodos cerrados

Este capítulo sobre raíces de ecuaciones se ocupa de métodos que aprovechan el hecho de que una función cambia de signo en la vecindad de una raíz. A estas técnicas se les llama *métodos cerrados*, o de *intervalos*, porque se necesita de dos valores iniciales para la raíz. Como su nombre lo indica, dichos valores iniciales deben "encerrar", o estar a ambos lados de la raíz. Los métodos particulares descritos aquí emplean diferentes estrategias para reducir sistemáticamente el tamaño del intervalo y así converger a la respuesta correcta.

Como preámbulo de estas técnicas se analizarán los métodos gráficos para representar tanto las funciones como sus raíces. Además de la utilidad de los métodos gráficos para determinar valores iniciales, también son útiles para visualizar las propiedades de las funciones y el comportamiento de los diversos métodos numéricos.

5.1 MÉTODOS GRÁFICOS

Un método simple para obtener una aproximación a la raíz de la ecuación $f(x) = 0$ consiste en graficar la función y observar dónde cruza el eje x. Este punto, que representa el valor de x para el cual $f(x) = 0$, ofrece una aproximación inicial de la raíz.

EJEMPLO 5.1 El método gráfico

Planteamiento del problema Utilice el método gráfico para determinar el coeficiente de resistencia c necesario para que un paracaidista de masa $m = 68.1$ kg tenga una velocidad de 40 m/s después de una caída libre de $t = 10$ s. *Nota:* La aceleración de la gravedad es de 9.81 m/s^2.

Solución Este problema se resuelve determinando la raíz de la ecuación (PT2.4) usando los parámetros $t = 10$, $g = 9.81$, $v = 40$ y $m = 68.1$:

$$f(c) = \frac{9.81(68.1)}{c}\left(1 - e^{-(c/68.1)10}\right) - 40$$

o

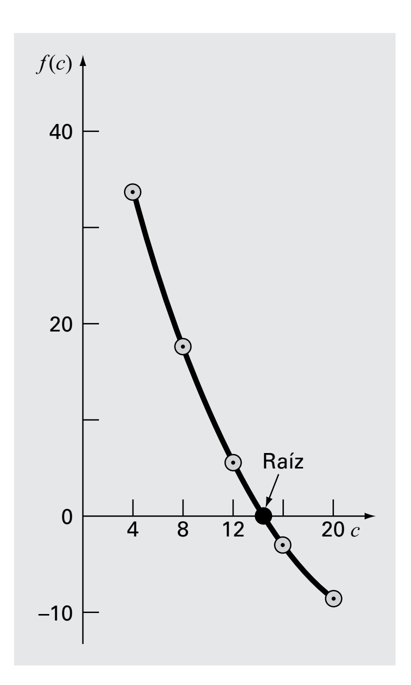
$$f(c) = \frac{668.06}{c}\left(1 - e^{-0.146843c}\right) - 40 \qquad \text{(E5.1.1)}$$

Diversos valores de c pueden sustituirse en el lado derecho de esta ecuación para calcular (ver tabla de la figura 5.1).

Estos puntos se grafican en la figura 5.1. La curva resultante cruza el eje c entre 12 y 16. Un vistazo a la gráfica proporciona una aproximación a la raíz de 14.75. La validez de la aproximación visual se verifica sustituyendo su valor en la ecuación (E5.1.1) para obtener

$$f(14.75) = \frac{668.06}{14.75}\left(1 - e^{-0.146843(14.75)}\right) - 40 = 0.100$$

que está cercano a cero. También se verifica por sustitución en la ecuación (PT2.3) junto con el valor de los parámetros de este ejemplo para dar



$$v = \frac{9.81(68.1)}{14.75}\left(1 - e^{-(14.75/68.1)10}\right) = 40.100$$

que es muy cercano a la velocidad de caída deseada de 40 m/s.

c	*f*(*c*)
4	34.190
8	17.712
12	6.114
16	–2.230
20	–8.368

Figura 5.1
El método gráfico para determinar las raíces de una ecuación.

Las técnicas gráficas tienen un valor práctico limitado, ya que no son precisas. Sin embargo, los métodos gráficos se utilizan para obtener aproximaciones de la raíz. Dichas aproximaciones se pueden usar como valores iniciales en los métodos numéricos analizados en este capítulo y en el siguiente.

Las interpretaciones gráficas, además de proporcionar estimaciones de la raíz, son herramientas importantes en la comprensión de las propiedades de las funciones y en la prevención de las fallas de los métodos numéricos. Por ejemplo, la figura 5.2 muestra algunas de las formas en las que la raíz puede encontrarse (o no encontrarse) en un intervalo definido por un límite inferior x_l y un límite superior x_u. La figura 5.2*b* representa el caso en que una sola raíz está acotada por los valores positivo y negativo de $f(x)$. Sin embargo, la figura 5.2*d*, donde $f(x_l)$ y $f(x_u)$ están también en lados opuestos del eje *x*, muestra tres raíces que se presentan en ese intervalo. En general, si $f(x_l)$ y $f(x_u)$ tienen signos opuestos, existe un número impar de raíces en el intervalo. Como se indica en las figuras 5.2*a* y *c*, si $f(x_l)$ y $f(x_u)$ tienen el mismo signo, no hay raíces o hay un número par de ellas entre los valores.

Aunque dichas generalizaciones son usualmente verdaderas, existen casos en que no se cumplen. Por ejemplo, las funciones tangenciales al eje *x* (figura 5.3*a*) y las funciones discontinuas (figura 5.3*b*) pueden violar estos principios. Un ejemplo de una función que es tangencial al eje *x* es la ecuación cúbica $f(x) = (x - 2)(x - 2)(x - 4)$. Observe que cuando $x = 2$, dos términos en este polinomio son iguales a cero. Matemáticamente, $x = 2$ se llama una ***raíz múltiple***. Al final del capítulo 6 se presentan técnicas que están diseñadas expresamente para localizar raíces múltiples.

La existencia de casos del tipo mostrado en la figura 5.3 dificulta el desarrollo de algoritmos generales para computadoras que garanticen la ubicación de todas las raíces en el intervalo. Sin embargo, cuando se usan los métodos expuestos en las siguientes secciones en conjunción con los métodos gráficos, resultan de gran utilidad para buscar muchas raíces en problemas de ecuaciones que se presentan rutinariamente en la ingeniería y en las matemáticas aplicadas.

EJEMPLO 5.2 Uso de gráficas por computadora para localizar raíces

Planteamiento del problema Las gráficas por computadora facilitan y mejoran la localización de las raíces de una ecuación. La función

$$f(x) = \text{sen } 10x + \cos 3x$$

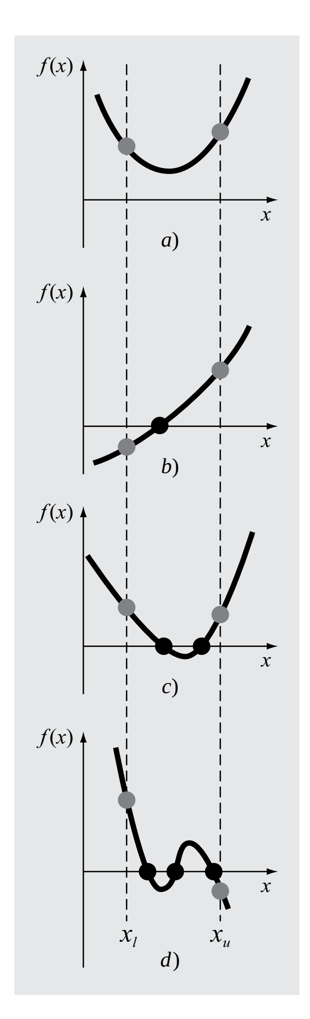


Figura 5.2
Ilustración de las formas generales en que puede ocurrir una raíz en un intervalo preescrito por los límites inferior x_l y superior x_u. Las figuras *a*) y *c*) muestran que si $f(x_l)$ y $f(x_u)$ tienen el mismo signo, entonces no habrá raíces dentro del intervalo o habrá un número par de ellas. Las figuras *b*) y *d*) muestran que si la función tiene signos diferentes en los puntos extremos, entonces habrá un número impar de raíces dentro del intervalo.

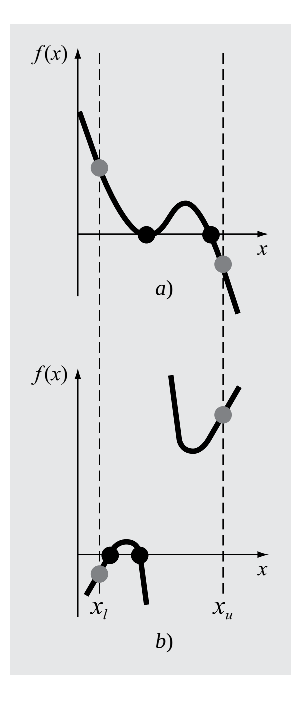


Figura 5.3
Ilustración de algunas excepciones a los casos generales mostrados en la figura 5.2. *a*) Pueden ocurrir raíces múltiples cuando la función es tangencial el eje *x*. En este caso, aunque los puntos extremos son de signos opuestos, hay un número par de intersecciones con el eje *x* en el intervalo. *b*) Función discontinua donde los puntos extremos de signo opuesto contienen un número par de raíces. Se requiere de estrategias especiales para determinar las raíces en estos casos.

tiene varias raíces en el rango que va de $x = 0$ a $x = 5$. Utilice gráficas por computadora para comprender mejor el comportamiento de esta función.

Solución Para generar gráficas se usan paquetes como Excel y MATLAB. En la figura 5.4*a* se presenta la gráfica de $f(x)$ desde $x = 0$ hasta $x = 5$. La gráfica muestra la existencia de varias raíces, incluyendo quizás una doble raíz alrededor de $x = 4.2$, donde $f(x)$ parece ser tangente al eje *x*. Se obtiene una descripción más detallada del comportamiento de $f(x)$ cambiando el rango de graficación, desde $x = 3$ hasta $x = 5$, como se muestra en la figura 5.4*b*. Finalmente, en la figura 5.4*c*, se reduce la escala vertical, de $f(x) = -0.15$ a $f(x) = 0.15$, y la escala horizontal se reduce, de $x = 4.2$ a $x = 4.3$. Esta gráfica muestra claramente que no existe una doble raíz en esta región y que, en efecto, hay dos raíces diferentes entre $x = 4.23$ y $x = 4.26$.

Las gráficas por computadora tienen gran utilidad en el estudio de los métodos numéricos. Esta posibilidad también puede tener muchas aplicaciones en otras materias de la escuela, así como en las actividades profesionales.

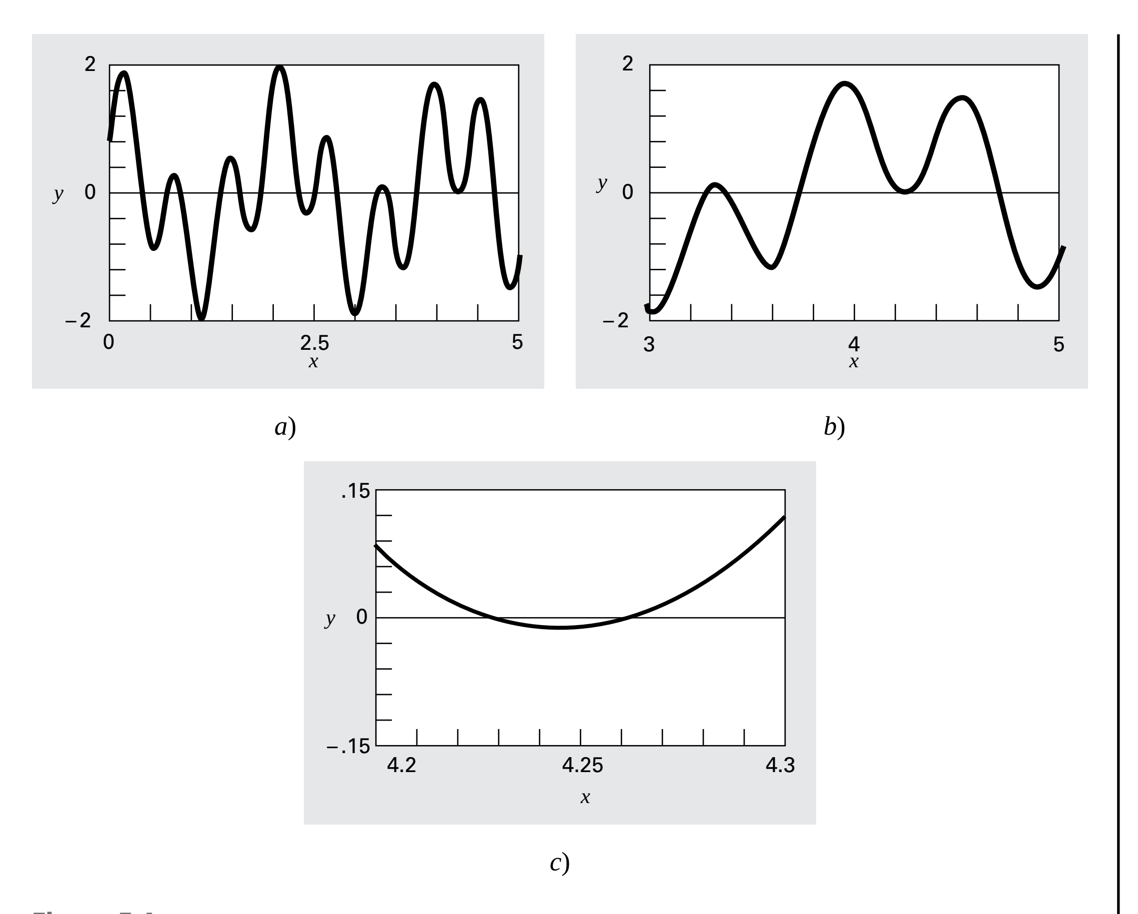


Figura 5.4
Amplificación progresiva de $f(x) = \text{sen } 10x + \cos 3x$ mediante la computadora. Estas gráficas interactivas le permiten al analista determinar que existen dos raíces distintas entre $x = 4.2$ y $x = 4.3$.

5.2 EL MÉTODO DE BISECCIÓN

Cuando se aplicaron las técnicas gráficas en el ejemplo 5.1, usted observó (figura 5.1) que $f(x)$ cambió de signo a ambos lados de la raíz. En general, si $f(x)$ es real y continúa en el intervalo que va desde x_l hasta x_u y $f(x_l)$ y $f(x_u)$ tienen signos opuestos, es decir,

$$f(x_l)\, f(x_u) < 0 \tag{5.1}$$

entonces hay al menos una raíz real entre x_l y x_u.

Los *métodos de búsqueda incremental* aprovechan esta característica localizando un intervalo en el que la función cambie de signo. Entonces, la localización del cambio de signo (y, en consecuencia, de la raíz) se logra con más exactitud al dividir el intervalo en varios subintervalos. Se investiga cada uno de estos subintervalos para encontrar el cambio de signo. El proceso se repite y la aproximación a la raíz mejora cada vez más en la medida que los subintervalos se dividen en intervalos cada vez más pequeños. Volveremos al tema de búsquedas incrementales en la sección 5.4.

El *método de bisección*, conocido también como de corte binario, de partición de intervalos o de Bolzano, es un tipo de búsqueda incremental en el que el intervalo se divide siempre a la mitad. Si la función cambia de signo sobre un intervalo, se evalúa el valor de la función en el punto medio. La posición de la raíz se determina situándola en el punto medio del subintervalo, dentro del cual ocurre un cambio de signo. El proceso se repite hasta obtener una mejor aproximación. En la figura 5.5 se presenta un algoritmo sencillo para los cálculos de la bisección. En la figura 5.6 se muestra una representación gráfica del método. Los siguientes ejemplos se harán a través de cálculos reales involucrados en el método.

Paso 1: Elija valores iniciales inferior, x_l, y superior, x_u, que encierren la raíz, de forma tal que la función cambie de signo en el intervalo. Esto se verifica comprobando que $f(x_l)\, f(x_u) < 0$.

Paso 2: Una aproximación de la raíz x_r se determina mediante:

$$x_r = \frac{x_l + x_u}{2}$$

Paso 3: Realice las siguientes evaluaciones para determinar en qué subintervalo está la raíz:

a) Si $f(x_l)f(x_r) < 0$, entonces la raíz se encuentra dentro del subintervalo inferior o izquierdo. Por lo tanto, haga $x_u = x_r$ y vuelva al paso 2.

b) Si $f(x_l)f(x_r) > 0$, entonces la raíz se encuentra dentro del subintervalo superior o derecho. Por lo tanto, haga $x_l = x_r$ y vuelva al paso 2.

c) Si $f(x_l)f(x_r) = 0$, la raíz es igual a x_r; termina el cálculo.

Figura 5.5

EJEMPLO 5.3 Bisección

Planteamiento del problema Emplee el método de bisección para resolver el mismo problema que se resolvió usando el método gráfico del ejemplo 5.1.

Solución El primer paso del método de bisección consiste en asignar dos valores iniciales a la incógnita (en este problema, c) que den valores de $f(c)$ con diferentes signos. En la figura 5.1 se observa que la función cambia de signo entre los valores 12 y 16. Por lo tanto, la estimación inicial de la raíz x_r se encontrará en el punto medio del intervalo

$$x_r = \frac{12+16}{2} = 14$$

Dicha aproximación representa un error relativo porcentual verdadero de $\varepsilon_t = 5.3\%$ (note que el valor verdadero de la raíz es 14.8011). A continuación calculamos el producto de los valores en la función en un límite inferior y en el punto medio:

$$f(12)f(14) = 6.114(1.611) = 9.850$$

que es mayor a cero y, por lo tanto, no ocurre cambio de signo entre el límite inferior y el punto medio. En consecuencia, la raíz debe estar localizada entre 14 y 16. Entonces, se crea un nuevo intervalo redefiniendo el límite inferior como 14 y determinando una nueva aproximación corregida de la raíz

$$x_r = \frac{14+16}{2} = 15$$

la cual representa un error porcentual verdadero $\varepsilon_t = 1.5\%$. Este proceso se repite para obtener una mejor aproximación. Por ejemplo,

$$f(14)f(15) = 1.611(-0.384) = -0.619$$

Por lo tanto, la raíz está entre 14 y 15. El límite superior se redefine como 15 y la raíz estimada para la tercera iteración se calcula así:

$$x_r = \frac{14+15}{2} = 14.5$$

que representa un error relativo porcentual $\varepsilon_t = 2.09\%$. Este método se repite hasta que el resultado sea suficientemente exacto para satisfacer sus necesidades.

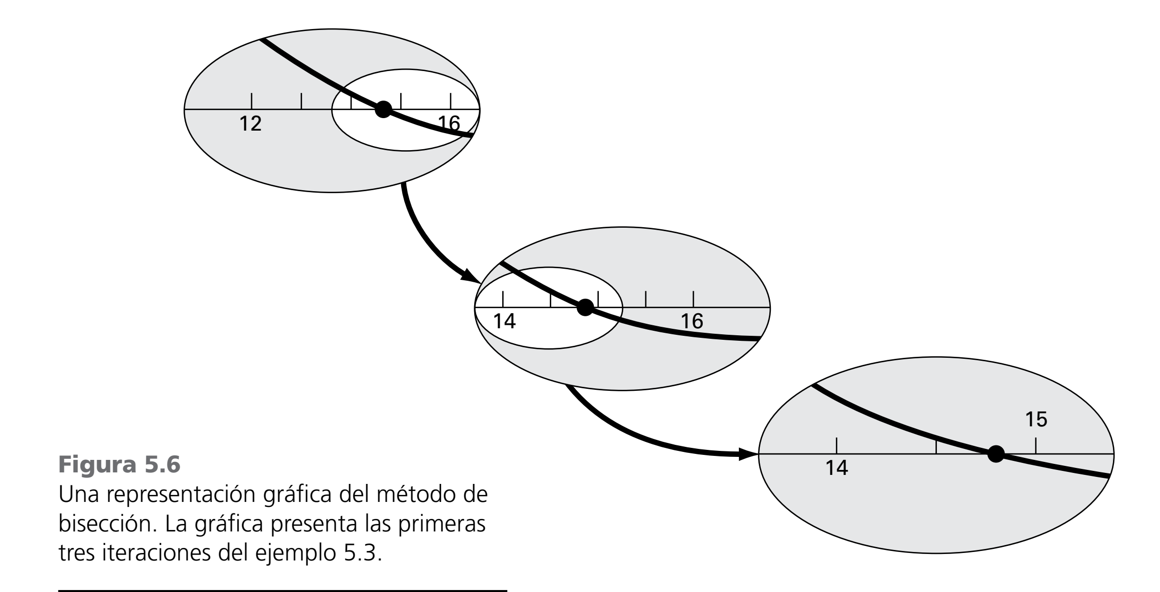


Figura 5.6
Una representación gráfica del método de bisección. La gráfica presenta las primeras tres iteraciones del ejemplo 5.3.

En el ejemplo anterior, se observa que el error verdadero no disminuye con cada iteración. Sin embargo, el intervalo donde se localiza la raíz se divide a la mitad en cada paso del proceso. Como se estudiará en la siguiente sección, el ancho del intervalo proporciona una estimación exacta del límite superior del error en el método de bisección.

5.2.1 Criterios de paro y estimaciones de errores

Terminamos el ejemplo 5.3 diciendo que el método se repite para obtener una aproximación más exacta de la raíz. Ahora debemos desarrollar un criterio objetivo para decidir cuándo debe terminar el método.

Una sugerencia inicial sería finalizar el cálculo cuando el error verdadero se encuentre por debajo de algún nivel prefijado. En el ejemplo 5.3 se observa que el error relativo baja de 5.3 a 2.0% durante el procedimiento de cálculo. Podemos decidir que el método termina cuando se alcance un error más bajo, por ejemplo, al 0.1%. Dicha estrategia es inconveniente, ya que la estimación del error en el ejemplo anterior se basó en el conocimiento del valor verdadero de la raíz de la función. Éste no es el caso de una situación real, ya que no habría motivo para utilizar el método si conocemos la raíz.

Por lo tanto, necesitamos estimar el error de forma tal que no se necesite el conocimiento previo de la raíz. Como se vio previamente en la sección 3.3, se puede calcular el error relativo porcentual ε_a de la siguiente manera [recuerde la ecuación (3.5)]:

$$\varepsilon_a = \left| \frac{x_r^{\text{nuevo}} - x_r^{\text{anterior}}}{x_r^{\text{nuevo}}} \right| 100\% \tag{5.2}$$

donde x_r^{nuevo} es la raíz en la iteración actual y x_r^{anterior} es el valor de la raíz en la iteración anterior. Se utiliza el valor absoluto, ya que por lo general importa sólo la magnitud de ε_a sin considerar su signo. Cuando ε_a es menor que un valor previamente fijado ε_s, termina el cálculo.

EJEMPLO 5.4 Estimación del error en la bisección

Planteamiento del problema Continúe con el ejemplo 5.3 hasta que el error aproximado sea menor que el criterio de terminación de $\varepsilon_s = 0.5\%$. Use la ecuación (5.2) para calcular los errores.

Solución Los resultados de las primeras dos iteraciones en el ejemplo 5.3 fueron 14 y 15. Al sustituir estos valores en la ecuación (5.2) se obtiene

$$|\varepsilon_a| = \left|\frac{15-14}{15}\right| 100\% = 6.667\,\%$$

Recuerde que el error relativo porcentual para la raíz estimada de 15 fue 1.3%. Por lo tanto, ε_a es mayor a ε_t. Este comportamiento se manifiesta en las otras iteraciones:

Iteración	x_l	x_u	x_r	ε_a (%)	ε_t (%)
1	12	16	14		5.413
2	14	16	15	6.667	1.344
3	14	15	14.5	3.448	2.035
4	14.5	15	14.75	1.695	0.345
5	14.75	15	14.875	0.840	0.499
6	14.75	14.875	14.8125	0.422	0.077

Así, después de seis iteraciones ε_a finalmente está por debajo de $\varepsilon_s = 0.5\%$, y el cálculo puede terminar.

Estos resultados se resumen en la figura 5.7. La naturaleza "desigual" del error verdadero se debe a que, en el método de la bisección, la raíz exacta se encuentra en cualquier lugar dentro del intervalo cerrado. Los errores verdadero y aproximado quedan distantes cuando el intervalo está centrado sobre la raíz verdadera. Ellos están cercanos cuando la raíz verdadera se halla en cualquier extremo del intervalo.

Aunque el error aproximado no proporciona una estimación exacta del error verdadero, la figura 5.7 sugiere que ε_a toma la tendencia general descendente de ε_t. Además, la gráfica muestra una característica muy interesante: que ε_a siempre es mayor que ε_t. Por lo tanto, cuando ε_a es menor que ε_s los cálculos se pueden terminar, con la confianza de saber que la raíz es al menos tan exacta como el nivel aceptable predeterminado.

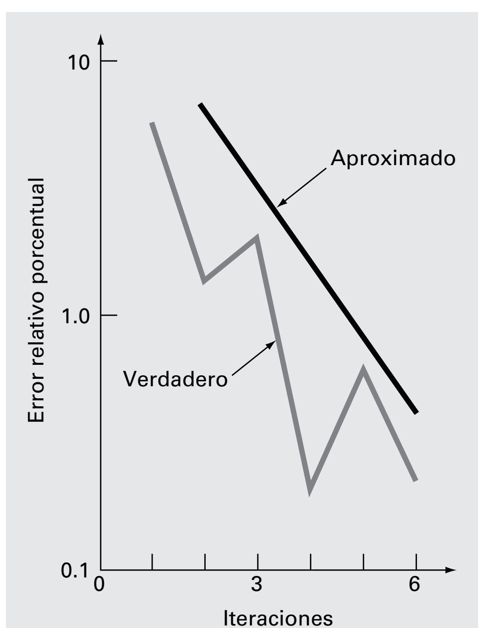


Figura 5.7
Errores en el método de bisección. Los errores verdadero y aproximado se grafican contra el número de iteraciones.

Aunque no es conveniente aventurar conclusiones generales a partir de un solo ejemplo, es posible demostrar que ε_a siempre será mayor que ε_t en el método de bisección. Esto se debe a que cada vez que se encuentra una aproximación a la raíz cuando se usan bisecciones como $x_r = (x_l + x_u)/2$, se sabe que la raíz verdadera se halla en algún lugar dentro del intervalo de $(x_u - x_l)/2 = \Delta x/2$. Por lo tanto, la raíz debe situarse dentro de $\pm\Delta x/2$ de la aproximación (figura 5.8). Así, cuando se terminó el ejemplo 5.3 podemos afirmar definitivamente que

$$x_r = 14.5 \pm 0.5$$

Debido a que $\Delta x/2 = x_r^{\text{nuevo}} - x_r^{\text{anterior}}$ (figura 5.9), la ecuación (5.2) proporciona un límite superior exacto del error verdadero. Para que se rebase este límite, la raíz verdadera tendría que estar fuera del intervalo que la contiene, lo cual, por definición, jamás ocurrirá en el método de bisección. El ejemplo 5.6 muestra otras técnicas de localización de raíces que no siempre resultan tan eficientes. Aunque el método de bisección por lo general es más lento que otros métodos, la claridad del análisis de error ciertamente es un aspecto positivo que puede volverlo atractivo para ciertas aplicaciones en ingeniería.

Antes de utilizar el programa de computadora para la bisección, debemos observar que las siguientes relaciones (figura 5.9)

$$x_r^{\text{nuevo}} - x_r^{\text{anterior}} = \frac{x_u - x_l}{2}$$

y

$$x_r^{\text{nuevo}} = \frac{x_l + x_u}{2}$$

pueden sustituirse en la ecuación (5.2) para desarrollar una formulación alternativa en la aproximación del error relativo porcentual

$$\varepsilon_a = \left|\frac{x_u - x_l}{x_u + x_l}\right| 100\% \tag{5.3}$$

Esta ecuación resulta idéntica a la ecuación (5.2) para la bisección. Además, permite calcular el error basándose en nuestros valores iniciales; es decir, en la primera iteración. Por ejemplo, en la primera iteración del ejemplo 5.2, el error aproximado se calcula como

$$\varepsilon_a = \left|\frac{16-12}{16+12}\right| 100\% = 14.29\%$$

Otro beneficio del método de bisección es que el número de iteraciones requerido para obtener un error absoluto se calcula *a priori*; esto es, antes de empezar las iteraciones, donde se observa que antes de empezar esta técnica, el error absoluto es

$$E_a^0 = x_u^0 - x_l^0 = \Delta x^0$$

donde los superíndices definen la iteración. Por lo tanto, antes de empezar el método se tiene la "iteración cero". Después de la primera iteración el error será

$$E_a^1 = \frac{\Delta x^0}{2}$$

Debido a que en cada iteración se reduce el error a la mitad, la fórmula general que relaciona el error y el número de iteraciones, n, es

$$E_a^n = \frac{\Delta x^0}{2^n} \tag{5.4}$$

Si $E_{a,d}$ es el error deseado, en esta ecuación se despeja

$$n = \frac{\log(\Delta x^0/E_{a,d})}{\log 2} = \log_2\left(\frac{\Delta x^0}{E_{a,d}}\right) \tag{5.5}$$

Probemos la fórmula. En el ejemplo 5.4, el intervalo inicial fue $\Delta x_0 = 16 - 12 = 4$. Después de seis iteraciones, el error absoluto era

$$E_a = \frac{|14.875 - 14.75|}{2} = 0.0625$$

Si se sustituyen esos valores en la ecuación (5.5) resulta

$$n = \frac{\log(4/0.0625)}{\log 2} = 6$$

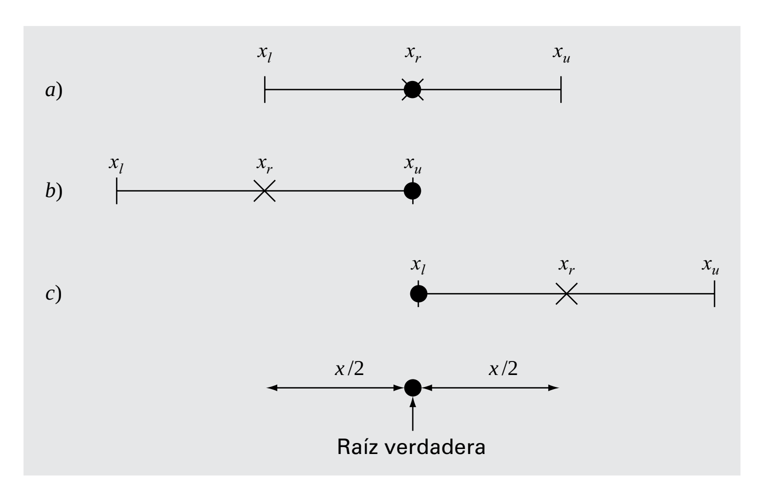


Figura 5.8
Tres formas en que un intervalo puede encerrar a la raíz. En *a*) el valor verdadero está en el centro del intervalo, mientras que en *b*) y *c*) el valor verdadero está cerca de los extremos. Observe que la diferencia entre el valor verdadero y el punto medio del intervalo jamás sobrepasa la longitud media del intervalo, o $\Delta x/2$.

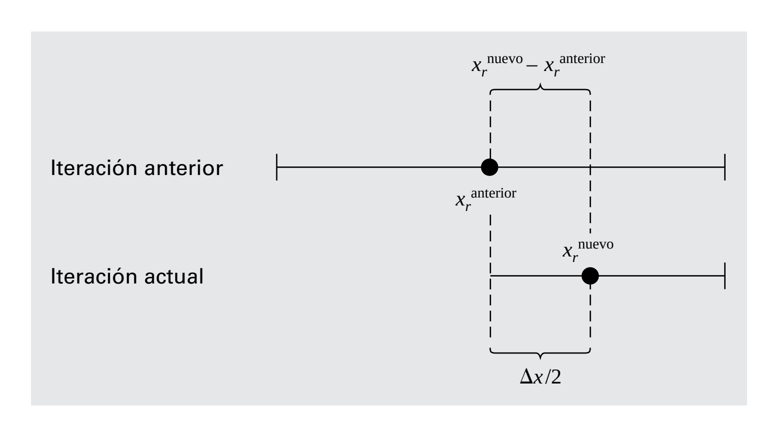


Figura 5.9
Representación gráfica de por qué la estimación del error para el método de bisección ($\Delta x/2$) es equivalente a la raíz estimada en la iteración actual (x_r^{nuevo}) menos la raíz aproximada en la iteración anterior (x_r^{anterior}).

Entonces, si se sabe de antemano que un error menor a 0.0625 es aceptable, la fórmula indica que con seis iteraciones se consigue el resultado deseado.

Aunque se ha puesto énfasis en el uso del error relativo por obvias razones, habrá casos (usualmente a través del conocimiento del contexto del problema) donde se podrá especificar el error ab-

soluto. En esos casos, la bisección junto con la ecuación (5.5) ofrece un útil algoritmo de localización de raíces. Se explorarán tales aplicaciones con los problemas al final del capítulo.

5.2.2 Algoritmo de bisección

El algoritmo en la figura 5.5 se extiende para incluir verificación del error (figura 5.10). El algoritmo emplea funciones definidas por el usuario para volver más eficientes la localización de las raíces y la evaluación de las funciones. Además, se le pone un límite superior al número de iteraciones. Por último, se incluye la verificación de errores para evitar la división entre cero durante la evaluación del error. Éste podría ser el caso cuando el intervalo está centrado en cero. En dicha situación la ecuación (5.2) tiende a infinito. Si esto ocurre, el programa saltará la evaluación de error en esa iteración.

El algoritmo en la figura 5.10 no es amigable al usuario; más bien está diseñado estrictamente para dar la respuesta. En el problema 5.14 al final del capítulo, se tendrá una tarea para volverlo fácil de usar y de entender.

5.2.3 Minimización de las evaluaciones de una función

El algoritmo de bisección de la figura 5.10 es adecuado si se quiere realizar la evaluación de una sola raíz de una función que es fácil de evaluar. Sin embargo, hay muchos casos en ingeniería que no son así. Por ejemplo, suponga que se quiere desarrollar un programa computacional que localice varias raíces. En tales casos, se tendría que llamar al algoritmo de la figura 5.10 miles o aun millones de veces en el transcurso de una sola ejecución.

Además, en un sentido más general, la función de una variable es tan sólo una entidad que regresa un solo valor para un solo valor que se le da. Visto de esta manera, las funciones no son simples fórmulas como las ecuaciones de una sola línea de código resueltas en los ejemplos anteriores de este capítulo. Por ejemplo, una función puede consistir de muchas líneas de código y su evaluación llega a tomar un tiempo importante de ejecución. En algunos casos, esta función incluso representaría un programa de computadora independiente.

Debido a ambos factores es imperativo que los algoritmos numéricos minimicen las evaluaciones de una función. A la luz de estas consideraciones, el algoritmo de la figura 5.10 es deficiente. En particular, observe que al hacer dos evaluaciones de una función por iteración, vuelve a calcular una de las funciones que se determinó en la iteración anterior.

La figura 5.11 proporciona un algoritmo modificado que no tiene esta deficiencia. Se han resaltado las líneas que difieren de la figura 5.10. En este caso, únicamente se calcula el valor de la nueva función para aproximar la raíz. Los valores calculados previamente son guardados y simplemente reasignados conforme el intervalo se reduce. Así, las $2n$ evaluaciones de la función se reducen a $n + 1$.

```
FUNCTION Bisect(xl, xu, es, imax, xr, iter, ea)
  iter = 0
  DO
    xrold = xr
    xr = (xl + xu)/2
    iter = iter + 1
    IF xr ≠ 0 THEN
       ea = ABS((xr - xrold) / xr) * 100
    END IF
    test = f(xl) * f(xr)
    IF test < 0 THEN
       xu = xr
    ELSE IF test > 0 THEN
       xl = xr
    ELSE
       ea = 0
    END IF
    IF ea < es OR iter ≥ imax EXIT
  END DO
  Bisect = xr
END Bisect
```

Figura 5.10
Pseudocódigo para la función que implementa el método de bisección.

5.3 MÉTODO DE LA FALSA POSICIÓN

Aun cuando la bisección es una técnica perfectamente válida para determinar raíces, su método de aproximación por "fuerza bruta" es relativamente ineficiente. La falsa posición es una alternativa basada en una visualización gráfica.

Un inconveniente del método de bisección es que al dividir el intervalo de x_l a x_u en intervalos iguales, no se toman en consideración las magnitudes de $f(x_l)$ y $f(x_u)$. Por ejemplo, si $f(x_l)$ está mucho más cercana a cero que $f(x_u)$, es lógico que la raíz se encuentre más cerca de x_l que de x_u (figura 5.12). Un método alternativo que aprovecha esta visualización gráfica consiste en unir $f(x_l)$ y $f(x_u)$ con una línea recta. La intersección de esta línea con el eje de las x representa una mejor aproximación de la raíz. El hecho de que se reemplace la curva por una línea recta da una “falsa posición” de la raíz; de aquí el nombre de *método de la falsa posición*, o en latín, *regula falsi*. También se le conoce como *método de interpolacion lineal*.

```
FUNCTION Bisect(xl, xu, es, imax, xr, iter, ea)
  iter = 0
  fl = f(xl)
  DO
    xrold = xr
    xr = (xl + xu) / 2
    fr = f(xr)
    iter = iter + 1
    IF xr ≠ 0 THEN
      ea = ABS((xr - xrold) / xr) * 100
    END IF
    test = fl * fr
    IF test < 0 THEN
      xu = xr
    ELSE IF test > 0 THEN
      xl = xr
      fl = fr
    ELSE
      ea = 0
    END IF
    IF ea < es OR iter ≥ imax EXIT
  END DO
  Bisect = xr
END Bisect
```

Figura 5.11
Pseudocódigo para el subprograma de bisección que minimiza las evaluaciones de la función.

Usando triángulos semejantes (figura 5.12), la intersección de la línea recta con el eje de las x se estima mediante

$$\frac{f(x_l)}{x_r - x_l} = \frac{f(x_u)}{x_r - x_u} \tag{5.6}$$

en la cual se despeja x_r (véase cuadro 5.1 para los detalles)

$$x_r = x_u - \frac{f(x_u)(x_l - x_u)}{f(x_l) - f(x_u)} \tag{5.7}$$

Ésta es la *fórmula de la falsa posición*. El valor de x_r calculado con la ecuación (5.7), reemplazará, después, a cualquiera de los dos valores iniciales, x_l o x_u, y da un valor de la función con el mismo signo de $f(x_r)$. De esta manera, los valores x_l y x_u siempre encierran la verdadera raíz. El proceso se repite hasta que la aproximación a la raíz sea adecuada. El algoritmo es idéntico al de la bisección (figura 5.5), excepto en que la ecuación (5.7) se usa en el paso 2. Además, se usa el mismo criterio de terminación [ecuación (5.2)] para concluir los cálculos.

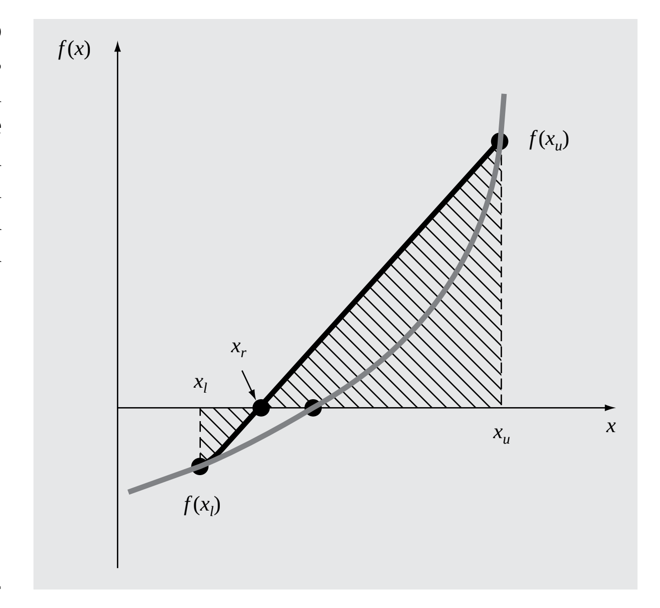


Figura 5.12
Representación gráfica del método de la falsa posición. Con los triángulos semejantes sombreados se obtiene la fórmula para el método.

Cuadro 5.1 Desarrollo del método de la falsa posición

Multiplicando en cruz la ecuación (5.6) obtenemos

$$f(x_l)(x_r - x_u) = f(x_u)(x_r - x_l)$$

Agrupando términos y reordenando:

$$x_r\,[f(x_l) - f(x_u)] = x_u\,f(x_l) - x_l\,f(x_u)$$

Dividiendo entre $f(x_l) - f(x_u)$:

$$x_r = \frac{x_u f(x_l) - x_l f(x_u)}{f(x_l) - f(x_u)} \qquad \text{(B5.1.1)}$$

Ésta es una de las formas del método de la falsa posición. Observe que permite el cálculo de la raíz x_r como una función de los valores iniciales inferior x_l y superior x_u. Ésta puede ponerse en una forma alternativa al separar los términos:

$$x_r = \frac{x_u f(x_l)}{f(x_l) - f(x_u)} - \frac{x_l f(x_u)}{f(x_l) - f(x_u)}$$

sumando y restando x_u en el lado derecho:

$$x_r = x_u + \frac{x_u f(x_l)}{f(x_l) - f(x_u)} - x_u - \frac{x_l f(x_u)}{f(x_l) - f(x_u)}$$

Agrupando términos se obtiene

$$x_r = x_u + \frac{x_u f(x_u)}{f(x_l) - f(x_u)} - \frac{x_l f(x_u)}{f(x_l) - f(x_u)}$$

o

$$x_r = x_u - \frac{f(x_u)(x_l - x_u)}{f(x_l) - f(x_u)}$$

la cual es la misma ecuación (5.7). Se utiliza esta forma porque implica una evaluación de la función y una multiplicación menos que la ecuación (B5.1.1). Además ésta es directamente comparable con el método de la secante, el cual se estudia en el capítulo 6.

EJEMPLO 5.5 Falsa posición

Planteamiento del problema Con el método de la falsa posición determine la raíz de la misma ecuación analizada en el ejemplo 5.1 [ecuación (E5.1.1)].

Solución Como en el ejemplo 5.3 se empieza el cálculo con los valores iniciales $x_l = 12$ y $x_u = 16$.

Primera iteración:

$x_l = 12 \qquad f(x_l) = 6.1139$

$x_u = 16 \qquad f(x_u) = -2.2303$

$$x_r = 16 - \frac{-2.2303(12 - 16)}{6.1139 - (-2.2303)} = 14.9309$$

que tiene un error relativo verdadero de 0.88 por ciento.

Segunda iteración:

$$f(x_l)\,f(x_r) = -1.5376$$

Por lo tanto, la raíz se encuentra en el primer subintervalo, y x_r se vuelve ahora el límite superior para la siguiente iteración, $x_u = 14.9309$:

$x_l = 12 \qquad f(x_l) = 6.1139$

$x_u = 14.9309 \qquad f(x_u) = -0.2515$

$$x_r = 14.9309 - \frac{-0.2515(12 - 14.9309)}{6.1139 - (-0.2515)} = 14.8151$$

el cual tiene error relativo verdadero y aproximado de 0.09 y 0.79 por ciento. Es posible realizar iteraciones adicionales para hacer una mejor aproximación de las raíces.

Se obtiene una idea más completa de la eficiencia de los métodos de bisección y de falsa posición al observar la figura 5.13, donde se muestra el error relativo porcentual verdadero de los ejemplos 5.4 y 5.5. Observe cómo el error decrece mucho más rápidamente en el método de la falsa posición que en el de la bisección, debido a un esquema más eficiente en el método de la falsa posición para la localización de raíces.

Recuerde que en el método de bisección el intervalo entre x_l y x_u se va haciendo más pequeño durante los cálculos. Por lo tanto, el intervalo, como se definió por $\Delta x/2 = |x_u - x_l|/2$ para la primera iteración, proporciona una medida del error en este método. Éste no es el caso con el método de la falsa posición, ya que uno de los valores iniciales puede permanecer fijo durante los cálculos, mientras que el otro converge hacia la raíz. Como en el caso del ejemplo 5.5, el extremo inferior x_l permanece en 12, mientras que x_u converge a la raíz. En tales casos, el intervalo no se acorta, sino que se aproxima a un valor constante.

El ejemplo 5.5 sugiere que la ecuación (5.2) representa un criterio de error muy conservador. De hecho, la ecuación (5.2) constituye una aproximación de la discrepancia en la iteración previa. Esto se debe a que para un caso, tal como el del ejemplo 5.5, donde el método converge rápidamente (por ejemplo, el error se va reduciendo casi un 100% de magnitud por cada iteración), la raíz para la iteración actual x_r^{nuevo} es una mejor aproximación al valor real de la raíz, que el resultado de la iteración previa x_r^{anterior}. Así, el numerador de la ecuación (5.2) representa la discrepancia de la iteración previa. En consecuencia, se nos asegura que al satisfacer la ecuación (5.2), la raíz se conocerá con mayor exactitud que la tolerancia preestablecida. Sin embargo, como se ve en la siguiente sección, existen casos donde el método de la falsa posición converge lentamente. En tales casos la ecuación (5.2) no es confiable y se debe desarrollar un criterio diferente de terminación.

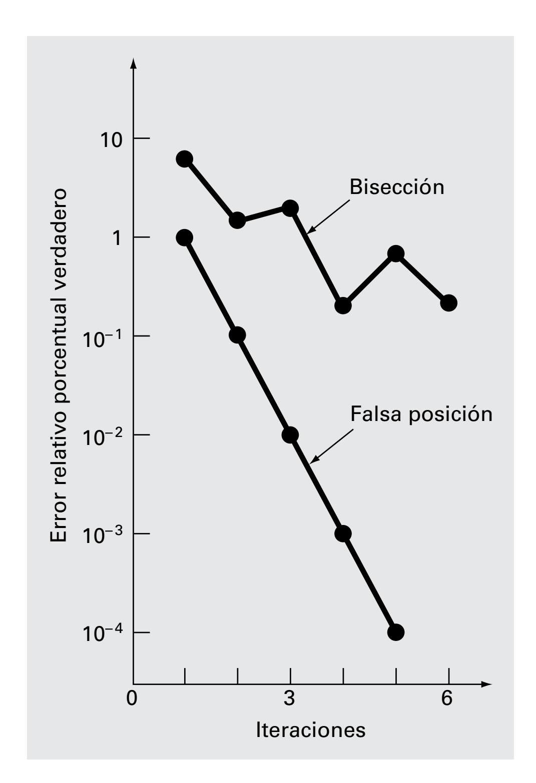


Figura 5.13
Comparación de los errores relativos de los métodos de bisección y de la falsa posición.

5.3.1 Desventajas del método de la falsa posición

Aunque el método de la falsa posición parecería ser siempre la mejor opción entre los métodos cerrados, hay casos donde funciona de manera deficiente. En efecto, como en el ejemplo siguiente, hay ciertos casos donde el método de bisección ofrece mejores resultados.

EJEMPLO 5.6 Un caso en el que la bisección es preferible a la falsa posición

Planteamiento del problema Con los métodos de bisección y de falsa posición localice la raíz de

$$f(x) = x^{10} - 1$$

entre $x = 0$ y 1.3.

Solución Usando bisección, los resultados se resumen como sigue

Iteración	x_l	x_u	x_r	ε_a (%)	ε_t (%)
1	0	1.3	0.65	100.0	35
2	0.65	1.3	0.975	33.3	2.5
3	0.975	1.3	1.1375	14.3	13.8
4	0.975	1.1375	1.05625	7.7	5.6
5	0.975	1.05625	1.015625	4.0	1.6

De esta manera, después de cinco iteraciones, el error verdadero se reduce a menos del 2%. Con la falsa posición se obtienen resultados muy diferentes:

Iteración	x_l	x_u	x_r	ε_a (%)	ε_t (%)
1	0	1.3	0.09430		90.6
2	0.09430	1.3	0.18176	48.1	81.8
3	0.18176	1.3	0.26287	30.9	73.7
4	0.26287	1.3	0.33811	22.3	66.2
5	0.33811	1.3	0.40788	17.1	59.2

Después de cinco iteraciones, el error verdadero sólo se ha reducido al 59%. Además, observe que $\varepsilon_a < \varepsilon_t$. Entonces, el error aproximado es engañoso. Se obtiene mayor claridad sobre estos resultados examinando una gráfica de la función. En la figura 5.14, la curva viola la premisa sobre la cual se basa la falsa posición; es decir, si $f(x_l)$ se encuentra mucho más cerca de cero que $f(x_u)$, la raíz se encuentra más cerca de x_l que de x_u (recuerde la figura 5.12). Sin embargo, debido a la forma de esta función ocurre lo contrario.

El ejemplo anterior ilustra que, por lo común, no es posible realizar generalizaciones con los métodos de obtención de raíces. Aunque un método como el de la falsa posición casi siempre es superior al de bisección, hay algunos casos que violan esta conclusión general. Por lo tanto, además de usar la ecuación (5.2), los resultados se deben verificar sustituyendo la raíz aproximada en la ecuación original y determinar si el resultado se acerca a cero. Esta prueba se debe incorporar en todos los programas que localizan raíces.

El ejemplo ilustra también una importante desventaja del método de la falsa posición: su unilateralidad. Es decir, conforme se avanza en las iteraciones, uno de los puntos limitantes del intervalo tiende a permanecer fijo. Esto puede llevar a una mala convergencia, especialmente en funciones con una curvatura importante. La sección siguiente ofrece una solución.

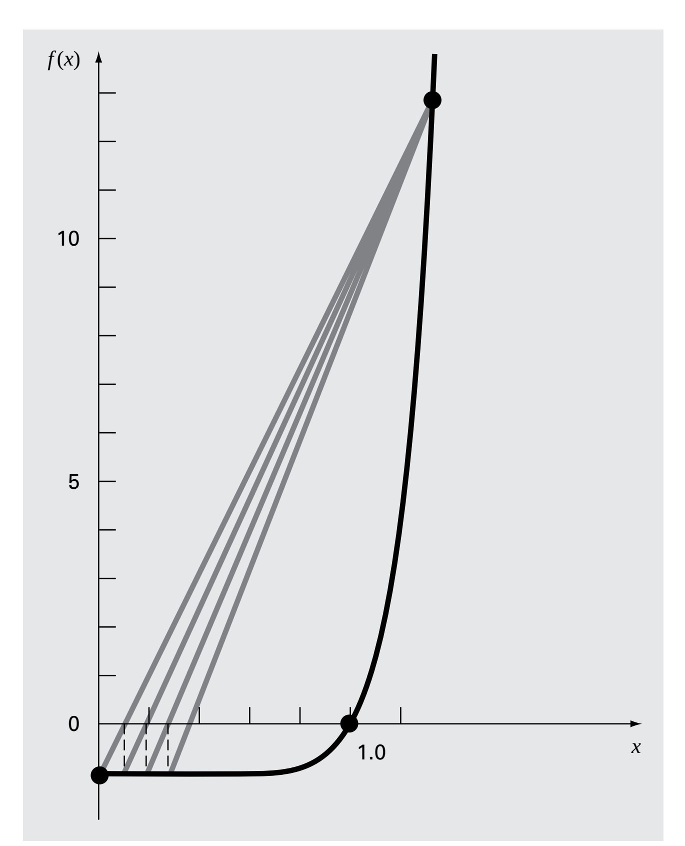


Figura 5.14
Gráfica de la función $f(x) = x^{10} - 1$, ilustrando la lentitud de convergencia del método de la falsa posición.

5.3.2 Falsa posición modificada

Una forma de disminuir la naturaleza unilateral de la falsa posición consiste en obtener un algoritmo que detecte cuando se "estanca" uno de los límites del intervalo. Si ocurre esto, se divide a la mitad el valor de la función en el punto de "estancamiento". A este método se le llama *método de la falsa posición modificado.*

El algoritmo dado en la figura 5.15 lleva a cabo dicha estrategia. Observe cómo se han usado contadores para determinar si uno de los límites del intervalo permanece fijo "estancado" durante dos iteraciones. Si ocurre así, el valor de la función en este valor de "estancamiento" se divide a la mitad.

La efectividad de este algoritmo se demuestra aplicándolo al ejemplo 5.6. Si se utiliza un criterio de terminación de 0.01% el método de bisección y el método estándar de falsa posición convergerán, respectivamente, después de 14 y 39 iteraciones. En cambio el método de la falsa posición modificado convergerá después de 12 iteraciones. De manera que para este ejemplo el método de la falsa posición modificado es más eficiente que el de bisección y muchísimo mejor que el método de la falsa posición no modificado.

5.4 BÚSQUEDAS POR INCREMENTOS Y DETERMINACIÓN DE VALORES INICIALES

Además de verificar una respuesta individual, se debe determinar si se han localizado todas las raíces posibles. Como se mencionó anteriormente, por lo general una gráfica de la función ayudará a realizar dicha tarea. Otra opción es incorporar una búsqueda incremental al inicio del programa. Esto consiste en empezar en un extremo del intervalo de interés y realizar evaluaciones de la función con pequeños incrementos a lo largo del intervalo. Si la función cambia de signo, se supone que la raíz está dentro del incremento. Los valores de x, al principio y al final del incremento, pueden servir como valores iniciales para una de las técnicas descritas en este capítulo.

Un problema potencial en los métodos de búsqueda por incremento es el de escoger la longitud del incremento. Si la longitud es muy pequeña, la búsqueda llega a consumir demasiado tiempo. Por otro lado, si la longitud es demasiado grande, existe la posibilidad de que raíces muy cercanas entre sí pasen inadvertidas (figura 5.16). El problema se complica con la posible existencia de raíces múltiples. Un remedio parcial para estos casos consiste en calcular la primera derivada de la función $f'(x)$ al inicio y al final de cada intervalo. Cuando la derivada cambia de signo, puede existir un máximo o un mínimo en ese intervalo, lo que sugiere una búsqueda más minuciosa para detectar la posibilidad de una raíz.

Aunque estas modificaciones o el empleo de un incremento muy fino ayudan a resolver el problema, se debe aclarar que métodos como el de la búsqueda incremental no siempre resultan sencillos. Será prudente complementar dichas técnicas automáticas con cualquier otra información que dé idea de la localización de las raíces. Esta información se puede encontrar graficando la función y entendiendo el problema físico de donde proviene la ecuación.

```
FUNCTION ModFalsePos(xl, xu, es, imax, xr, iter, ea)
  iter = 0
  fl = f(xl)
  fu = f(xu)
  DO
    xrold = xr
    xr = xu - fu * (xl - xu) / (fl - fu)
    fr = f(xr)
    iter = iter + 1
    IF xr <> 0 THEN
      ea = Abs((xr - xrold) / xr) * 100
    END IF
    test = fl * fr
    IF test < 0 THEN
      xu = xr
      fu = f(xu)
      iu = 0
      il = il +1
      If il ≥ 2 THEN fl = fl / 2
    ELSE IF test > 0 THEN
      xl = xr
      fl = f (xl)
      il = 0
      iu = iu + 1
      IF iu ≥ 2 THEN fu = fu / 2
    ELSE
      ea = 0
    END IF
    IF ea < es OR iter ≥ imax THEN EXIT
  END DO
  ModFalsePos = xr
END ModFalsePos
```

Figura 5.15
Pseudocódigo para el método de la falsa posición modificado.

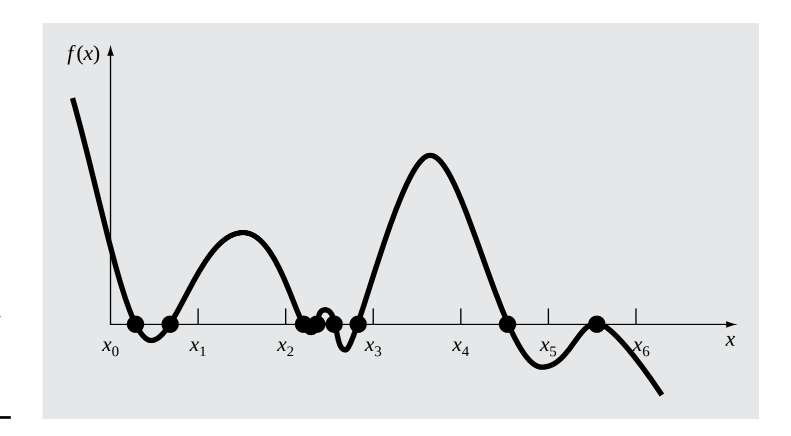


Figura 5.16
Casos donde las raíces pueden pasar inadvertidas debido a que la longitud del incremento en el método de búsqueda incremental es demasiado grande. Observe que la última raíz a la derecha es múltiple y podría dejar de considerarse independientemente de la longitud del incremento.

PROBLEMAS

5.1 Determine las raíces reales de $f(x) = -0.5x^2 + 2.5x + 4.5$:
a) Gráficamente.
b) Empleando la fórmula cuadrática.
c) Usando el método de bisección con tres iteraciones para determinar la raíz más grande. Emplee como valores iniciales $x_l = 5$ y $x_u = 10$. Calcule el error estimado ε_a y el error verdadero ε_t para cada iteración.

5.2 Determine las raíces reales de $f(x) = 5x^3 - 5x^2 + 6x - 2$:
a) Gráficamente.
b) Utilizando el método de bisección para localizar la raíz. Use los valores iniciales $x_l = 0$ y $x_u = 1$ iterando hasta que el error estimado ε_a se encuentre debajo de $\varepsilon_s = 10\%$.

5.3 Determine las raíces reales de $f(x) = -25 + 82x - 90x^2 + 44x^3 - 8x^4 + 0.7x^5$:
a) Gráficamente.
b) Usando el método de bisección para localizar la raíz más grande con $\varepsilon_s = 10\%$. Utilice como valores iniciales $x_l = 0.5$ y $x_u = 1.0$.
c) Realice el mismo cálculo que en *b*), pero con el método de la falsa posición y $\varepsilon_s = 0.2\%$.

5.4 *a*) Determine gráficamente las raíces de $f(x) = -12 - 21x + 18x^2 - 2.75x^3$. Además, determine la primera raíz de la función con *b*) bisección y *c*) posición falsa. Para *b*) y *c*), utilice valores iniciales para $x_l = -1$ y $x_u = 0$, y un criterio de detención de 1%.

5.5 Localice la primera raíz no trivial de sen $x = x^2$, donde x está en radianes. Use una técnica gráfica y bisección con un intervalo inicial de 0.5 a 1. Haga el cálculo hasta que ε_a sea menor que $\varepsilon_s = 2\%$. Realice también una prueba de error sustituyendo la respuesta final en la ecuación original.

5.6 Determine la raíz real de $\ln(x^2) = 0.7$:
a) Gráficamente.
b) Empleando tres iteraciones en el método de bisección con los valores iniciales $x_l = 0.5$ y $x_u = 2$.
c) Usando tres iteraciones del método de la falsa posición, con los mismos valores iniciales de *b*).

5.7 Determine la raíz real de $f(x) = (0.8 - 0.3x)/x$:
a) Analíticamente.
b) Gráficamente.
c) Empleando tres iteraciones en el método de la falsa posición, con valores iniciales de 1 a 3, calcule el error aproximado ε_a y el error verdadero ε_t en cada iteración. ¿Hay algún problema con el resultado?

5.8 Calcule la raíz cuadrada positiva de 18 usando el método de la falsa posición con $\varepsilon_s = 0.5\%$. Emplee como valores iniciales $x_l = 4$ y $x_u = 5$.

5.9 Encuentre la raíz positiva más pequeña de la función (x está en radianes) $x^2|\cos\sqrt{x}| = 5$ usando el método de la falsa posición. Para localizar el intervalo en donde se encuentra la raíz, grafique primero esta función para valores de x entre 0 y 5. Realice el cálculo hasta que ε_a sea menor que $\varepsilon_s = 1\%$. Compruebe su respuesta final sustituyéndola en la función original.

5.10 Encuentre la raíz positiva de $f(x) = x^4 - 8x^3 - 35x^2 + 450x - 1\,001$, utilizando el método de la falsa posición. Tome como valores iniciales a $x_l = 4.5$ y $x_u = 6$, y ejecute cinco iteraciones. Calcule los errores tanto aproximado como verdadero, con base en el hecho de que la raíz es 5.60979. Emplee una gráfica para explicar sus resultados y hacer el cálculo dentro de un $\varepsilon_s = 1.0\%$.

5.11 Determine la raíz real de $x^{3.5} = 80$:
a) En forma analítica.
b) Con el método de la falsa posición dentro de $\varepsilon_s = 2.5\%$. Haga elecciones iniciales de 2.0 a 5.0.

5.12 Dada

$$f(x) = -2x^6 - 1.5x^4 + 10x + 2$$

Use el método de la bisección para determinar el *máximo* de esta función. Haga elecciones iniciales de $x_l = 0$ y $x_u = 1$, y realice iteraciones hasta que el error relativo aproximado sea menor que 5%.

5.13 La velocidad υ de un paracaidista que cae está dada por

$$\upsilon = \frac{gm}{c}\left(1 - e^{-(c/m)t}\right)$$

donde $g = 9.81$ m/s^2. Para un paracaidista con coeficiente de resistencia de $c = 15$ kg/s, calcule la masa m de modo que la velocidad sea $\upsilon = 36$ m/s en $t = 10$ s. Utilice el método de la falsa posición para determinar m a un nivel de $\varepsilon_s = 0.1\%$.

5.14 Use bisección para determinar el coeficiente de resistencia necesario para que un paracaidista de 82 kg tenga una velocidad de 36 m/s después de 4 s de caída libre. *Nota:* La aceleración de la gravedad es 9.81 m/s^2. Comience con valores iniciales de $x_l = 3$ y $x_u = 5$. Itere hasta que el error relativo aproximado caiga por debajo de 2%. Realice también una detección de errores sustituyendo su respuesta final en la ecuación original.

5.15 Como se ilustra en la figura P5.15, la velocidad del agua, υ (m/s), en la descarga de un tanque cilíndrico a través de un tubo largo se puede calcular como

$$\upsilon = \sqrt{2gH}\,\tanh\left(\frac{\sqrt{2gH}}{2L}t\right)$$

donde $g = 9.81$ m/s^2, H = carga hidrostática inicial (m), L = longitud de tubo (m) y t = tiempo transcurrido (s). Determine la carga hidrostática necesaria para obtener $\upsilon = 5$ m/s en 2.5 s para un tubo de 4 m de longitud *a*) gráficamente, *b*) por bisección y *c*) con posición falsa. Utilice los valores iniciales de $x_1 = 0$ y $x_u = 2$ m, con un criterio de detención de $\varepsilon_s = 1\%$. Revise sus resultados.

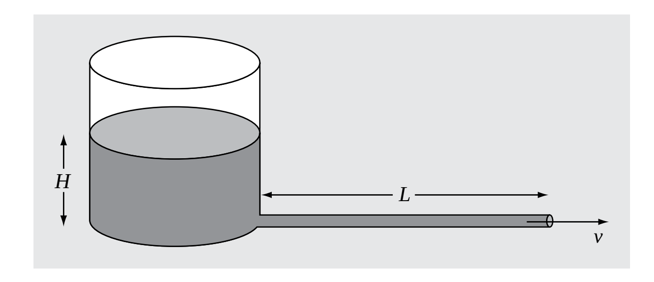


Figura P5.15

5.16 Por un canal trapezoidal fluye agua a una tasa de $Q = 20$ m^3/s. La profundidad crítica y para dicho canal satisface la ecuación

$$0 = 1 - \frac{Q^2}{gA_c^3}B$$

donde $g = 9.81\text{m/s}^2$, A_c = área de la sección transversal (m^2), y B = ancho del canal en la superficie (m). Para este caso, el ancho y el área de la sección transversal se relacionan con la profundidad y por medio de

$$B = 3 + y \qquad \text{y} \qquad A_c = 3y + \frac{y^2}{2}$$

Resuelva para la profundidad crítica con el uso de los métodos *a*) gráfico, *b*) bisección y *c*) falsa posición. En los incisos *b*) y *c*), haga elecciones iniciales de $x_l = 0.5$ y $x_u = 2.5$, y ejecute iteraciones hasta que el error aproximado caiga por debajo del 1% o el número de interaciones supere a 10. Analice sus resultados.

5.17 Suponga el lector que está diseñando un tanque esférico (véase la figura P5.17) para almacenar agua para un poblado pequeño en un país en desarrollo. El volumen de líquido que puede contener se calcula con

$$V = \pi h^2 \frac{[3R - h]}{3}$$

donde V = volumen [m^3], h = profundidad del agua en el tanque [m] y R = radio del tanque [m].

Si R = 3 m, ¿a qué profundidad debe llenarse el tanque de modo que contenga 30 m^3? Haga tres iteraciones con el método de la falsa posición a fin de obtener la respuesta. Determine el error relativo aproximado después de cada iteración. Utilice valores iniciales de 0 y R.

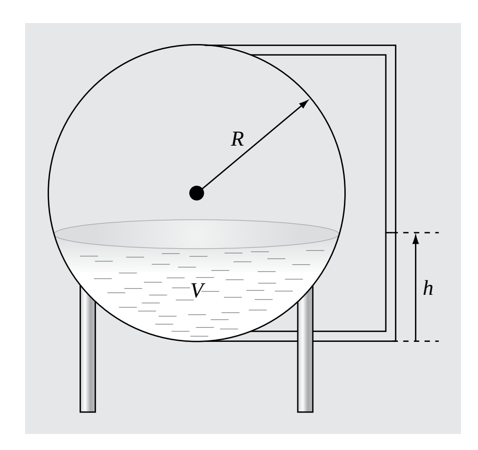


Figura P5.17

5.18 La concentración de saturación de oxígeno disuelto en agua dulce se calcula con la ecuación (APHA, 1992)

$$\ln o_{sf} = -139.34411 + \frac{1.575701 \times 10^5}{T_a} - \frac{6.642308 \times 10^7}{T_a^2} + \frac{1.243800 \times 10^{10}}{T_a^3} - \frac{8.621949 \times 10^{11}}{T_a^4}$$

donde o_{sf} = concentración de saturación de oxígeno disuelto en agua dulce a 1 atm (mg/L) y T_a = temperatura absoluta (K). Recuerde el lector que $T_a = T + 273.15$, donde T = temperatura (°C). De acuerdo con esta ecuación, la saturación disminuye con el incremento de la temperatura. Para aguas naturales comunes en climas templados, la ecuación se usa para determinar que la concentración de oxígeno varía de 14.621 mg/L a 0°C a 6.413 mg/L a 40°C. Dado un valor de concentración de oxígeno, puede emplearse esta fórmula y el método de bisección para resolver para la termperatura en °C.

a) Si los valores iniciales son de 0 y 40°C, con el método de la bisección, ¿cuántas iteraciones se requerirían para determinar la temperatura con un error absoluto de 0.05°C?

b) Desarrolle y pruebe un programa para el método de bisección a fin de determinar T como función de una concentración dada de oxígeno, con un error absoluto preespecificado como en el inciso *a*). Dadas elecciones iniciales de 0 y 40°C, pruebe su programa para un error absoluto de 0.05°C para los casos siguientes: o_{sf} = 8, 10 y 12 mg/L. Compruebe sus resultados.

5.19 De acuerdo con el *principio de Arquímedes*, la fuerza de *flotación* es igual al peso de fluido desplazado por la porción sumergida de un objeto. Para la esfera ilustrada en la figura P5.19, use la bisección para determinar la altura h de la porción que queda encima del agua. Utilice los siguientes valores para su cálculo: r = 1 m, ρ_s = densidad de la esfera = 200 kg/m^3 y ρ_w = densidad del agua = 1 000 kg/m^3. Observe que el volumen de la porción de la esfera por encima del agua se puede calcular mediante

$$V = \frac{\pi h^2}{3}(3r - h)$$

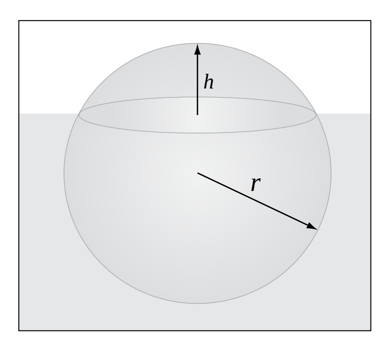


Figura P5.19

5.20 Realice el mismo cálculo que en el problema 5.19, pero para el *frústum* de un cono, como se ilustra en la figura P5.20. Use los siguientes valores para su cálculo: r_1 = 0.5 m, r_2 = 1 m, h = 1 m, ρ_f = densidad del frústum = 200 kg/m^3, y ρ_w = densidad del agua = 1 000 kg/m^3. Observe que el volumen del frústum está dado por

$$V = \frac{\pi h}{3}(r_1^2 + r_2^2 + r_1 r_2)$$

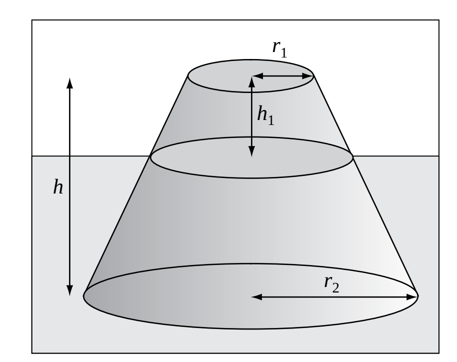


Figura P5.20

5.21 Integre el algoritmo que se bosquejó en la figura 5.10, en forma de subprograma completo para el método de bisección amigable para el usuario. Entre otras cosas:

a) Construya enunciados de documentación en el subprograma a fin de identificar lo que se pretende que realice cada sección.
b) Etiquete la entrada y la salida.
c) Agregue una comprobación de la respuesta, en la que se sustituya la estimación de la raíz en la función original para verificar si el resultado final se acerca a cero.
d) Pruebe el subprograma por medio de repetir los cálculos de los ejemplos 5.3 y 5.4.

5.22 Desarrolle un subprograma para el método de bisección que minimice las evaluaciones de la función, con base en el pseudocódigo que se presenta en la figura 5.11. Determine el número de evaluaciones de la función (n) para el total de iteraciones. Pruebe el programa con la repetición del ejemplo 5.6.

5.23 Desarrolle un programa amigable para el usuario para el método de la falsa posición. La estructura del programa debe ser similar al algoritmo de la bisección que se bosquejó en la figura 5.10. Pruebe el programa con la repetición del ejemplo 5.5.

5.24 Desarrolle un subprograma para el método de la falsa posición que minimice las evaluaciones de la función en forma similar a la figura 5.11. Determine el número de evaluaciones de la función (n) para el total de iteraciones. Pruebe el programa por medio de la duplicación del ejemplo 5.6.

5.25 Desarrolle un subprograma amigable para el usuario para el método de la falsa posición modificado, con base en la figura 5.15. Pruebe el programa con la determinación de la raíz de la función del ejemplo 5.6. Ejecute corridas hasta que el error relativo porcentual verdadero esté por debajo de 0.01%. Elabore una gráfica en un entorno semilogarítmico de los errores relativo, porcentual, aproximado y verdadero, *versus* el número de iteraciones. Interprete los resultados.

5.26 Desarrolle una función para bisección de modo similar a la figura 5.10. Sin embargo, en vez de usar iteraciones máximas y la ecuación (5.2), utilice la ecuación (5.5) como criterio de detención. Asegúrese de redondear el resultado de la ecuación (5.5) al máximo entero siguiente. Pruebe su función resolviendo el ejemplo 5.3 usando $E_{a,d}$ = 0.0001.

CAPÍTULO 6

Métodos abiertos

En los métodos cerrados del capítulo anterior la raíz se encuentra dentro de un intervalo predeterminado por un límite inferior y otro superior. La aplicación repetida de estos métodos siempre genera aproximaciones cada vez más cercanas al valor verdadero de la raíz. Se dice que tales métodos son *convergentes* porque se acercan progresivamente a la raíz a medida que se avanza en el cálculo (figura 6.1*a*).

En contraste, los *métodos abiertos* descritos en este capítulo se basan en fórmulas que requieren únicamente de un solo valor de inicio x o que empiecen con un par de ellos, pero que no necesariamente encierran la raíz. Éstos, algunas veces, *divergen* o se alejan de la raíz verdadera a medida que se avanza en el cálculo (figura 6.1*b*). Sin embargo, cuando los métodos abiertos convergen (figura 6.1*c*), en general lo hacen mucho más rápido que los métodos cerrados. Empecemos el análisis de los métodos abiertos con una versión simple que es útil para ilustrar su forma general y también para demostrar el concepto de convergencia.

6.1 ITERACIÓN SIMPLE DE PUNTO FIJO

Como se dijo antes, los métodos abiertos emplean una fórmula para predecir la raíz. Esta fórmula puede desarrollarse como una *iteración simple de punto fijo* (también llamada iteración de un punto o sustitución sucesiva o método de punto fijo), al arreglar la ecuación $f(x) = 0$ de tal modo que x esté del lado izquierdo de la ecuación:

$$x = g(x) \tag{6.1}$$

Esta transformación se realiza mediante operaciones algebraicas o simplemente sumando x a cada lado de la ecuación original. Por ejemplo,

$$x^2 - 2x + 3 = 0$$

se arregla para obtener

$$x = \frac{x^2 + 3}{2}$$

mientras que sen $x = 0$ puede transformarse en la forma de la ecuación (6.1) sumando x a ambos lados para obtener

$$x = \text{sen } x + x$$

La utilidad de la ecuación (6.1) es que proporciona una fórmula para predecir un nuevo valor de x en función del valor anterior de x. De esta manera, dado un valor inicial para la raíz x_i,

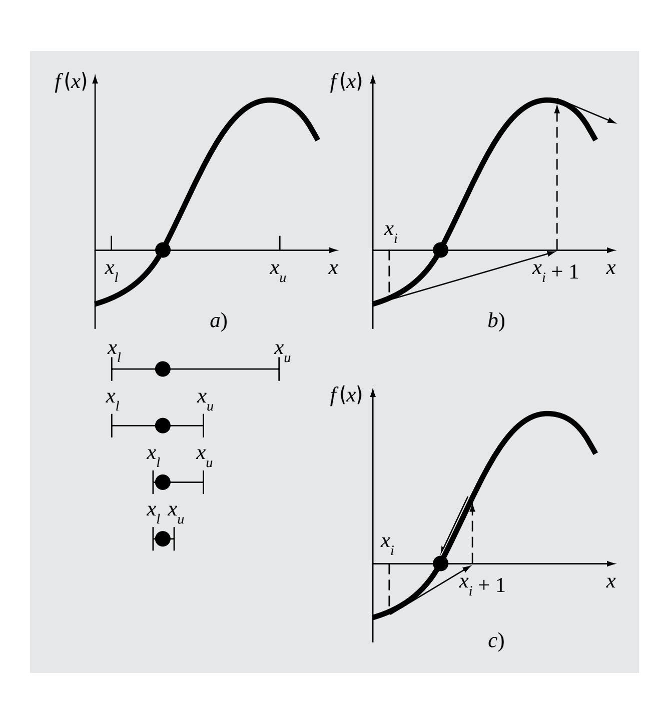


Figura 6.1
Representación gráfica de las diferencias fundamentales entre los métodos *a*) cerrados, *b*) y *c*) los métodos abiertos para el cálculo de raíces. En *a*) se ilustra el método de bisección, donde la raíz está contenida dentro del intervalo dado por x_l y x_u. En contraste, en los métodos abiertos, ilustrados en *b*) y *c*), se utiliza una fórmula para dirigirse de x_i a x_{i+1}, con un esquema iterativo. Así, el método puede *b*) diverger o *c*) converger rápidamente, dependiendo de los valores iniciales.

la ecuación (6.1) se utiliza para obtener una nueva aproximación x_{i+1}, expresada por la fórmula iterativa

$$x_{i+1} = g(x_i) \tag{6.2}$$

Como en otras fórmulas iterativas de este libro, el error aproximado de esta ecuación se calcula usando el error normalizado estimador [ecuación (3.5)]:

$$\varepsilon_a = \left| \frac{x_{i+1} - x_i}{x_{i+1}} \right| 100\%$$

EJEMPLO 6.1 Iteración simple de punto fijo

Planteamiento del problema Use una iteración simple de punto fijo para localizar la raíz de $f(x) = e^{-x} - x$.

Solución La función se puede separar directamente y expresarse en la forma de la ecuación (6.2) como

$$x_{i+1} = e^{-x_i}$$

Empezando con un valor inicial $x_0 = 0$, se aplica esta ecuación iterativa para calcular

i	x_i	ε_a (%)	ε_t (%)
0	0		100.0
1	1.000000	100.0	76.3
2	0.367879	171.8	35.1
3	0.692201	46.9	22.1
4	0.500473	38.3	11.8
5	0.606244	17.4	6.89
6	0.545396	11.2	3.83
7	0.579612	5.90	2.20
8	0.560115	3.48	1.24
9	0.571143	1.93	0.705
10	0.564879	1.11	0.399

De esta manera, se puede observar que cada iteración acerca cada vez más el valor aproximado al valor verdadero de la raíz: 0.56714329.

6.1.1 Convergencia

Note que el error relativo porcentual verdadero en cada iteración del ejemplo 6.1 es más o menos proporcional (por un factor de 0.5 a 0.6) al error de la iteración anterior. Esta propiedad, conocida como *convergencia lineal*, es característica de la iteración de punto fijo.

Además de la “velocidad” de convergencia, en este momento debemos enfatizar la “posibilidad” de convergencia. Los conceptos de convergencia y divergencia se pueden ilustrar gráficamente. Recuerde que en la sección 5.1 se graficó una función para visualizar su estructura y comportamiento (ejemplo 5.1). Ese método se emplea en la figura 6.2*a* para la función $f(x) = e^{-x} - x$. Un método gráfico alternativo consiste en separar la ecuación en dos partes, de esta manera

$$f_1(x) = f_2(x)$$

Entonces las dos ecuaciones

$$y_1 = f_1(x) \tag{6.3}$$

y

$$y_2 = f_2(x) \tag{6.4}$$

se grafican por separado (figura 6.2*b*). Así, los valores de x correspondientes a las intersecciones de estas dos funciones representan las raíces de $f(x) = 0$.

EJEMPLO 6.2 El método gráfico de las dos curvas

Planteamiento del problema Separe la ecuación $e^{-x} - x = 0$ en dos partes y determine su raíz en forma gráfica.

Solución Reformule la ecuación como $y_1 = x$ y $y_2 = e^{-x}$. Al tabular las funciones se obtienen los siguientes valores:

x	y_1	y_2
0.0	0.0	1.000
0.2	0.2	0.819
0.4	0.4	0.670
0.6	0.6	0.549
0.8	0.8	0.449
1.0	1.0	0.368

Estos puntos se grafican en la figura 6.2*b*. La intersección de las dos curvas indica una raíz estimada de aproximadamente $x = 0.57$, que corresponde al valor donde la curva de la figura 6.2*a* cruza el eje x.

El método de las dos curvas también se utiliza para ilustrar la convergencia y divergencia de la iteración de punto fijo. En primer lugar, la ecuación (6.1) se reexpresa como un par de ecuaciones $y_1 = x$ y $y_2 = g(x)$. Estas dos ecuaciones se grafican por separado. Como en el caso de las ecuaciones (6.3) y (6.4), las raíces de $f(x) = 0$ corresponden al valor de la abscisa en la intersección de las dos curvas. En la figura 6.3 se grafican la función $y_1 = x$ y cuatro formas diferentes de la función $y_2 = g(x)$.

En el primer caso (figura 6.3*a*), el valor inicial x_0 sirve para determinar el punto $[x_0, g(x_0)]$ correspondiente a la curva y_2. El punto (x_1, x_1) se encuentra moviéndose horizontalmente a la izquierda hasta la curva y_1. Estos movimientos son el equivalente a la primera iteración en el método de punto fijo:

$$x_1 = g(x_0)$$

De esta manera, tanto en la ecuación como en la gráfica se usa un valor inicial x_0 para obtener una aproximación de x_1. La siguiente iteración consiste en moverse al punto $[x_1, g(x_1)]$ y después a (x_2, x_2). Esta iteración es equivalente a la ecuación:

$$x_2 = g(x_1)$$

La solución en la figura 6.3*a* es *convergente*, ya que la aproximación de x se acerca más a la raíz con cada iteración. Lo mismo ocurre en la figura 6.3*b*. Sin embargo, éste no es el caso en las figuras 6.3*c* y 6.3*d*, donde las iteraciones divergen de la raíz. Observe que la convergencia ocurre únicamente cuando el valor absoluto de la pendiente de $y_2 = g(x)$ es menor al valor de la pendiente de $y_1 = x$, es decir, cuando $|g'(x)| < 1$. En el cuadro 6.1 se presenta un desarrollo teórico de este resultado.

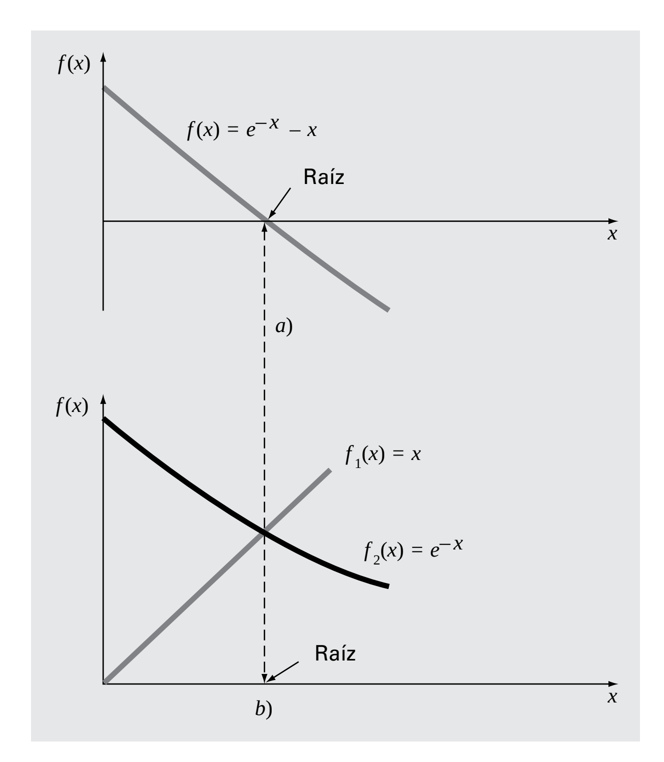


Figura 6.2
Dos métodos gráficos para determinar la raíz de $f(x) = e^{-x} - x$. *a*) La raíz como un punto donde la función cruza el eje x; *b*) la raíz como la intersección de las dos funciones componentes.

Cuadro 6.1 Convergencia de la iteración de punto fijo

Al analizar la figura 6.3, se debe notar que la iteración de punto fijo converge si, en la región de interés, $|g'(x)| < 1$. En otras palabras, la convergencia ocurre si la magnitud de la pendiente de $g(x)$ es menor que la pendiente de la recta $f(x) = x$. Esta observación puede demostrarse teóricamente. Recuerde que la ecuación iterativa es

$$x_{i+1} = g(x_i)$$

Suponga que la solución verdadera es

$$x_r = g(x_r)$$

Al restar estas dos ecuaciones se obtiene

$$x_r - x_{i+1} = g(x_r) - g(x_i) \qquad \text{(C6.1.1)}$$

El *teorema del valor medio de la derivada* (recuerde la sección 4.1.1) establece que si una función $g(x)$ y su primera derivada son continuas en un intervalo $a \le x \le b$, entonces existe al menos un valor de $x = \xi$ dentro del intervalo para el que

$$g'(\xi) = \frac{g(b) - g(a)}{b - a} \qquad \text{(C6.1.2)}$$

El lado derecho de esta ecuación es la pendiente de la recta que une a $g(a)$ y $g(b)$. Así, el teorema del valor medio establece que existe al menos un punto entre a y b que tiene una pendiente, denotada por $g'(\xi)$, que es paralela a la línea que une $g(a)$ con $g(b)$ (recuerde la figura 4.3).

Ahora, si se hace $a = x_i$ y $b = x_r$, el lado derecho de la ecuación (C6.1.1) se expresa como

$$g(x_r) - g(x_i) = (x_r - x_i)g'(\xi)$$

donde ξ se encuentra en alguna parte entre x_i y x_r. Este resultado se sustituye en la ecuación (C6.1.1) para obtener

$$x_r - x_{i+1} = (x_r - x_i)g'(\xi) \qquad \text{(C6.1.3)}$$

Si el error verdadero en la iteración i se define como

$$E_{t,i} = x_r - x_i$$

entonces la ecuación (C6.1.3) se convierte en

$$E_{t,i+1} = g'(\xi)E_{t,i}$$

En consecuencia, si $|g'(x)| < 1$, entonces los errores disminuyen con cada iteración. Si $|g'(x)| > 1$, los errores crecen. Observe también que si la derivada es positiva, los errores serán positivos y, por lo tanto, la solución iterativa será monótona (figuras 6.3*a* y 6.3*c*). Si la derivada es negativa, entonces los errores oscilarán (figuras 6.3*b* y 6.3*d*).

Un corolario de este análisis establece que cuando el método converge, el error es proporcional y menor que el error en la iteración anterior. Por tal razón se dice que la iteración simple de punto fijo es *linealmente convergente*.

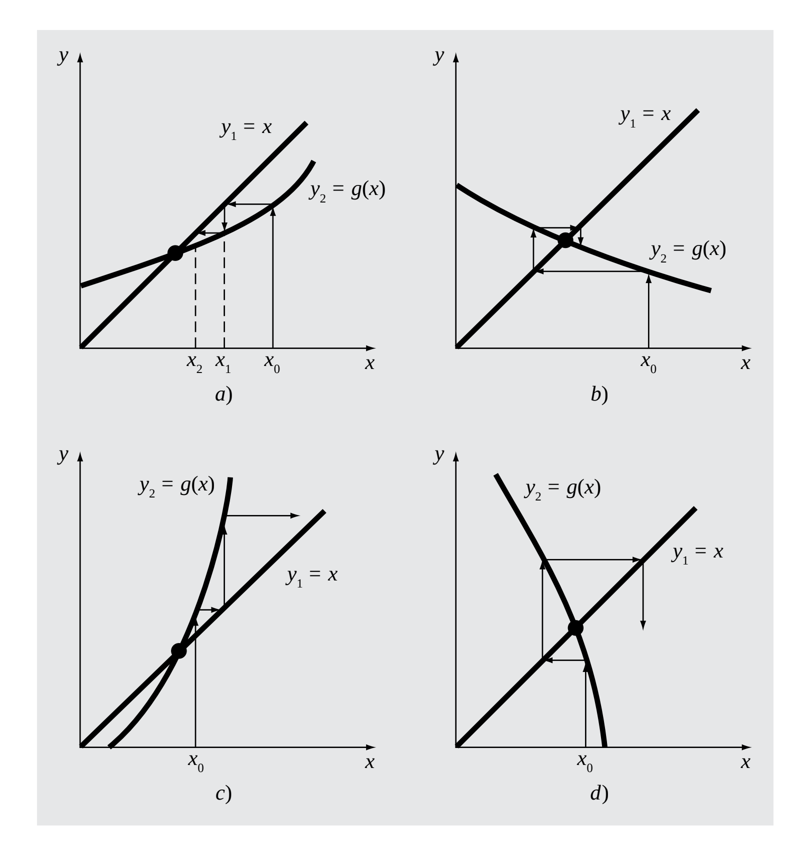


Figura 6.3
Redes de iteración que describen la convergencia (*a* y *b*) y la divergencia (*c* y *d*) de la iteración simple de punto fijo. Las gráficas *a*) y *c*) tienen un comportamiento monótono; mientras que *b*) y *d*) tienen un comportamiento oscilatorio o en espiral. Deberá notar que la convergencia se obtiene cuando $|g'(x)| < 1$.

6.1.2 Algoritmo para la iteración de punto fijo

El algoritmo para la iteración de punto fijo es simple en extremo. Consta de un *loop* o ciclo que calcula en forma iterativa nuevas aproximaciones hasta satisfacer el criterio de terminación. En la figura 6.4 se muestra el pseudocódigo para el algoritmo. Se pueden programar de manera similar otros métodos abiertos, la modificación principal consiste en cambiar la fórmula iterativa que se utiliza para calcular la nueva raíz.

```
FUNCTION Fixpt(x0, es, imax iter, ea)
  xr = x0
  iter = 0
  DO
    xrold = xr
    xr = g(xrold)
    iter = iter + 1
    IF xr ≠ 0 THEN
      ea = |(xr - xrold)/xr|·100
    END IF
    IF ea < es OR iter ≥ imax EXIT
  END DO
  Fixpt = xr
END Fixpt
```

Figura 6.4
Pseudocódigo para la iteración de punto fijo. Note que otros métodos abiertos pueden diseñarse en este formato general.

6.2 MÉTODO DE NEWTON-RAPHSON

Tal vez, de las fórmulas para localizar raíces, la fórmula de Newton-Raphson (figura 6.5) sea la más ampliamente utilizada. Si el valor inicial para la raíz es x_i, entonces se puede trazar una tangente desde el punto $[x_i, f(x_i)]$. Por lo común, el punto donde esta tangente cruza al eje x representa una aproximación mejorada de la raíz.

El método de Newton-Raphson se deduce a partir de esta interpretación geométrica (un método alternativo basado en la serie de Taylor se describe en el cuadro 6.2). De la figura 6.5, se tiene que la primera derivada en x es equivalente a la pendiente:

$$f'(x_i) = \frac{f(x_i) - 0}{x_i - x_{i+1}} \tag{6.5}$$

que se arregla para obtener

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \tag{6.6}$$

la cual se conoce como *fórmula de Newton-Raphson*.

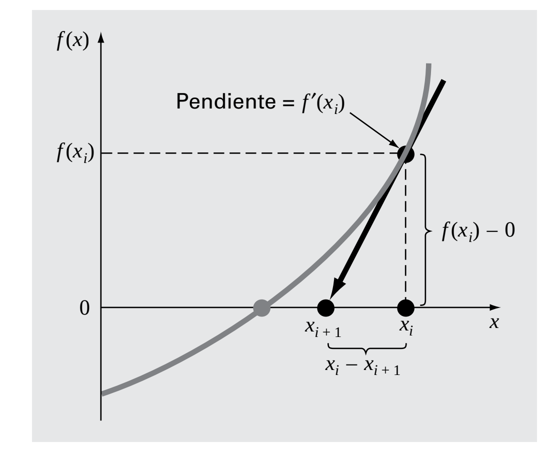


Figura 6.5
Representación gráfica del método de Newton-Raphson. Se extrapola una tangente a la función en x_i [esto es, $f'(x_i)$] hasta el eje x para obtener una estimación de la raíz en x_{i+1}.

EJEMPLO 6.3 Método de Newton-Raphson

Planteamiento del problema Utilice el método de Newton-Raphson para calcular la raíz de $f(x) = e^{-x} - x$ empleando como valor inicial $x_0 = 0$.

Solución La primera derivada de la función es

$$f'(x) = -e^{-x} - 1$$

que se sustituye, junto con la función original en la ecuación (6.6), para tener

$$x_{i+1} = x_i - \frac{e^{-x_i} - x_i}{-e^{-x_i} - 1}$$

Empezando con un valor inicial $x_0 = 0$, se aplica esta ecuación iterativa para calcular

i	x_i	ε_t (%)
0	0	100
1	0.500000000	11.8
2	0.566311003	0.147
3	0.567143165	0.0000220
4	0.567143290	$< 10^{-8}$

Así, el método converge rápidamente a la raíz verdadera. Observe que el error relativo porcentual verdadero en cada iteración disminuye mucho más rápido que con la iteración simple de punto fijo (compare con el ejemplo 6.1).

6.2.1 Criterio de terminación y estimación de errores

Como en los otros métodos para localizar raíces, la ecuación (3.5) se utiliza como un criterio de terminación. No obstante, el desarrollo del método con base en la serie de Taylor (cuadro 6.2) proporciona una comprensión teórica respecto a la velocidad de convergencia expresada por $E_{i+1} = O(E_i^2)$. De esta forma, el error debe ser más o menos proporcional al cuadrado del error anterior. En otras palabras, el número de cifras significativas de precisión aproximadamente se duplica en cada iteración. Dicho comportamiento se examina en el siguiente ejemplo.

EJEMPLO 6.4 Análisis de error en el método de Newton-Raphson

Planteamiento del problema Como se dedujo del cuadro 6.2, el método de Newton-Raphson es convergente en forma cuadrática. Es decir, el error es más o menos proporcional al cuadrado del error anterior:

$$E_{t,\,i+1} \cong \frac{-f''(x_r)}{2f'(x_r)} E_{t,i}^2 \tag{E6.4.1}$$

Examine esta fórmula y observe si concuerda con los resultados del ejemplo 6.3.

Solución La primera derivada de $f(x) = e^{-x} - x$ es

$$f'(x) = -e^{-x} - 1$$

que se evalúa en $x_r = 0.56714329$ para dar $f'(0.56714329) = -1.56714329$. La segunda derivada es:

$$f''(x) = e^{-x}$$

la cual se evalúa como $f''(0.56714329) = 0.56714329$. Estos resultados se sustituyen en la ecuación (E6.4.1):

$$E_{t,i+1} \cong -\frac{0.56714329}{2(-1.56714329)} E_{t,i}^2 = 0.18095 E_{t,i}^2$$

En el ejemplo 6.3, el error inicial fue $E_{t,0} = 0.56714329$, el cual se sustituye en la ecuación de error que predice

$$E_{t,1} \cong 0.18095(0.56714329)^2 = 0.0582$$

que es cercano al error verdadero de 0.06714329. En la siguiente iteración,

$$E_{t,2} \cong 0.18095(0.06714329)^2 = 0.0008158$$

que también se compara de manera favorable con el error verdadero 0.0008323. Para la tercera iteración,

$$E_{t,3} \cong 0.18095(0.0008323)^2 = 0.000000125$$

que es el error obtenido en el ejemplo 6.3. Así, la estimación del error mejora, ya que conforme nos acercamos a la raíz, x y ξ se aproximan mejor mediante x_r [recuerde nuestra suposición al ir de la ecuación (C6.2.5) a la ecuación (C6.2.6) en el cuadro 6.2]. Finalmente:

$$E_{t,4} \cong 0.18095(0.000000125)^2 = 2.83 \times 10^{-15}$$

Así, este ejemplo ilustra que el error en el método de Newton-Raphson para este caso es, de hecho, proporcional (por un factor de 0.18095) al cuadrado del error en la iteración anterior.

Cuadro 6.2 Deducción y análisis del error del método de Newton-Raphson

Además de la deducción geométrica [ecuaciones (6.5) y (6.6)], el método de Newton-Raphson también se desarrolla a partir de la expansión de la serie de Taylor. Esta deducción alternativa es muy útil en el sentido de que provee cierta comprensión sobre la velocidad de convergencia del método.

Recuerde del capítulo 4 que la expansión de la serie de Taylor se puede representar como

$$f(x_{i+1}) = f(x_i) + f'(x_i)(x_{i+1} - x_i) + \frac{f''(\xi)}{2!}(x_{i+1} - x_i)^2 \quad \text{(C6.2.1)}$$

donde ξ se encuentra en alguna parte del intervalo desde x_i hasta x_{i+1}. Truncando la serie de Taylor después del término de la primera derivada, se obtiene una versión aproximada:

$$f(x_{i+1}) \cong f(x_i) + f'(x_i)(x_{i+1} - x_i)$$

En la intersección con el eje x, $f(x_{i+1})$ debe ser igual a cero, o

$$0 = f(x_i) + f'(x_i)(x_{i+1} - x_i) \quad \text{(C6.2.2)}$$

de donde se puede despejar

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

que es idéntica a la ecuación (6.6). De esta forma, se ha deducido la fórmula de Newton-Raphson usando una serie de Taylor.

Además de este desarrollo, la serie de Taylor sirve para estimar el error de la fórmula. Esto se logra observando que si se utilizan todos los términos de la serie de Taylor se obtendrá un resultado exacto. En tal situación $x_{i+1} = x_r$, donde x es el valor verdadero de la raíz. Sustituyendo este valor junto con $f(x_r) = 0$ en la ecuación (C6.2.1) se obtiene

$$0 = f(x_i) + f'(x_i)(x_r - x_i) + \frac{f''(\xi)}{2!}(x_r - x_i)^2 \quad \text{(C6.2.3)}$$

La ecuación (C6.2.2) se resta de la ecuación (C6.2.3) para obtener

$$0 = f'(x_i)(x_r - x_{i+1}) + \frac{f''(\xi)}{2!}(x_r - x_i)^2 \quad \text{(C6.2.4)}$$

Ahora, observe que el error es igual a la diferencia entre x_{i+1} y el valor verdadero x_r, como en

$$E_{t,\,i+1} = x_r - x_{i+1}$$

y la ecuación (C6.2.4) se expresa como

$$0 = f'(x_i)E_{t,\,i+1} + \frac{f''(\xi)}{2!}E_{t,i}^2 \quad \text{(C6.2.5)}$$

Si se supone que hay convergencia, entonces tanto x_i como ξ se deberán aproximar a la raíz x_r y la ecuación (C6.2.5) se reordena para obtener

$$E_{t,\,i+1} = \frac{-f''(x_r)}{2f'(x_r)}E_{t,i}^2 \quad \text{(C6.2.6)}$$

De acuerdo con la ecuación (C6.2.6), el error es proporcional al cuadrado del error anterior. Esto significa que el número de cifras decimales correctas aproximadamente se duplica en cada iteración. A este comportamiento se le llama *convergencia cuadrática*. El ejemplo 6.4 ilustra esta propiedad.

6.2.2 Desventajas del método de Newton-Raphson

Aunque en general el método de Newton-Raphson es muy eficiente, hay situaciones donde se comporta de manera deficiente. Por ejemplo, en el caso especial de raíces múltiples que se analizará más adelante en este capítulo. Sin embargo, también cuando se trata de raíces simples, se encuentran dificultades, como en el siguiente ejemplo.

EJEMPLO 6.5 Ejemplo de una función que converge lentamente con el método de Newton-Raphson

Planteamiento del problema Determine la raíz positiva de $f(x) = x^{10} - 1$ usando el método de Newton-Raphson y un valor inicial $x = 0.5$.

Solución La fórmula de Newton-Raphson en este caso es:

$$x_{i+1} = x_i - \frac{x_i^{10} - 1}{10x_i^9}$$

que se utiliza para calcular:

Iteración	x
0	0.5
1	51.65
2	46.485
3	41.8365
4	37.65285
5	33.887565
.	
.	
.	
∞	1.0000000

De esta forma, después de la primera predicción deficiente, la técnica converge a la raíz verdadera, 1, pero muy lentamente.

Además de la convergencia lenta debido a la naturaleza de la función, es posible que se presenten otras dificultades, como se ilustra en la figura 6.6. Por ejemplo, la figura 6.6*a* muestra el caso donde un punto de inflexión [esto es, $f''(x) = 0$] ocurre en la vecindad de una raíz. Observe que las iteraciones que empiezan con x_0 divergen progresivamente de la raíz. En la figura 6.6*b* se ilustra la tendencia del método de Newton-Raphson a oscilar alrededor de un mínimo o máximo local. Tales oscilaciones pueden persistir o, como en la figura 6.6*b*, alcanzar una pendiente cercana a cero, después de lo cual la solución se aleja del área de interés. En la figura 6.6*c* se muestra cómo un valor inicial cercano a una raíz salta a una posición varias raíces más lejos. Esta tendencia a alejarse del área de interés se debe a que se encuentran pendientes cercanas a cero. En efecto, una pendiente cero [$f'(x) = 0$] es un verdadero desastre, ya que causa una división entre cero en la fórmula de Newton-Raphson [ecuación (6.6)]. En forma gráfica (figura 6.6*d*), esto significa que la solución se dispara horizontalmente y jamás toca al eje *x*.

De manera que no hay un criterio general de convergencia para el método de Newton-Raphson. Su convergencia depende de la naturaleza de la función y de la exactitud del valor inicial. La única solución en estos casos es tener un valor inicial que sea “suficientemente” cercano a la raíz. ¡Y para algunas funciones ningún valor inicial funcionará! Los buenos valores iniciales por lo común se predicen con un conocimiento del problema físico o mediante el uso de recursos alternativos, como las gráficas, que proporcionan mayor claridad en el comportamiento de la solución. Ante la falta de un criterio general de convergencia se sugiere el diseño de programas computacionales eficientes que reconozcan la convergencia lenta o la divergencia. La siguiente sección está enfocada hacia dichos temas.

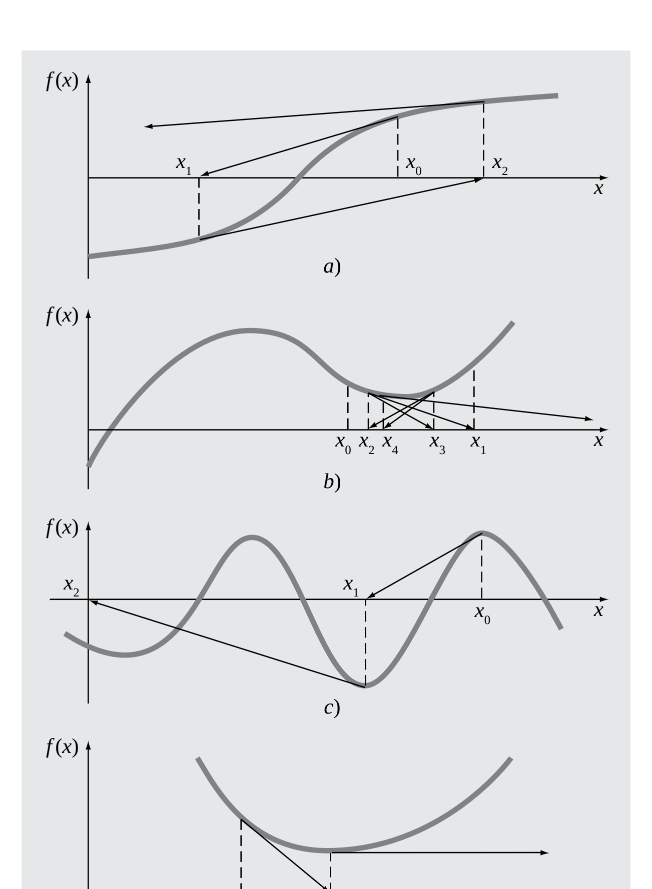


Figura 6.6
Cuatro casos donde el método de Newton-Raphson exhibe una convergencia deficiente.

6.2.3 Algoritmo para el método de Newton-Raphson

Un algoritmo para el método de Newton-Raphson se obtiene fácilmente al sustituir la ecuación (6.6) por la fórmula predictiva [ecuación (6.2)] en la figura 6.4. Observe, sin embargo, que el programa también debe modificarse para calcular la primera derivada. Esto se logra incluyendo simplemente una función definida por el usuario.

Además, a la luz del análisis anterior sobre los problemas potenciales del método de Newton-Raphson, el programa se podría mejorar incorporando algunas consideraciones adicionales:

1. Se debe incluir una rutina de graficación en el programa.
2. Al final de los cálculos, se necesitará sustituir siempre la raíz final calculada en la función original, para determinar si el resultado se acerca a cero. Esta prueba protege el desarrollo del programa contra aquellos casos en los que se presenta convergencia lenta u oscilatoria, la cual puede llevar a valores pequeños de ε_a, mientras que la solución aún está muy lejos de una raíz.
3. El programa deberá incluir siempre un límite máximo permitido del número de iteraciones para estar prevenidos contra soluciones oscilantes, de lenta convergencia o divergentes que podrían persistir en forma interminable.
4. El programa deberá alertar al usuario para que tome en cuenta la posibilidad de que $f'(x)$ sea igual a cero en cualquier momento durante el cálculo.

6.3 EL MÉTODO DE LA SECANTE

Un problema potencial en la implementación del método de Newton-Raphson es la evaluación de la derivada. Aunque esto no es un inconveniente para los polinomios ni para muchas otras funciones, existen algunas funciones cuyas derivadas en ocasiones resultan muy difíciles de calcular. En dichos casos, la derivada se puede aproximar mediante una diferencia finita dividida hacia atrás, como en (figura 6.7)

$$f'(x_i) \cong \frac{f(x_{i-1}) - f(x_i)}{x_{i-1} - x_i}$$

Esta aproximación se sustituye en la ecuación (6.6) para obtener la siguiente ecuación iterativa:

$$x_{i+1} = x_i - \frac{f(x_i)(x_{i-1} - x_i)}{f(x_{i-1}) - f(x_i)} \qquad (6.7)$$

La ecuación (6.7) es la fórmula para el *método de la secante*. Observe que el método requiere de dos valores iniciales de x. Sin embargo, debido a que no se necesita que $f(x)$ cambie de signo entre los valores dados, este método no se clasifica como un método cerrado.

EJEMPLO 6.6 El método de la secante

Planteamiento del problema Con el método de la secante calcule la raíz de $f(x) = e^{-x} - x$. Comience con los valores iniciales $x_{-1} = 0$ y $x_0 = 1.0$.

Solución Recuerde que la raíz real es 0.56714329…

Primera iteración:

$x_{-1} = 0 \qquad f(x_{-1}) = 1.00000$

$x_0 = 1 \qquad f(x_0) = -0.63212$

$$x_1 = 1 - \frac{-0.63212(0-1)}{1-(-0.63212)} = 0.61270 \qquad \varepsilon_t = 8.0\%$$

Segunda iteración:

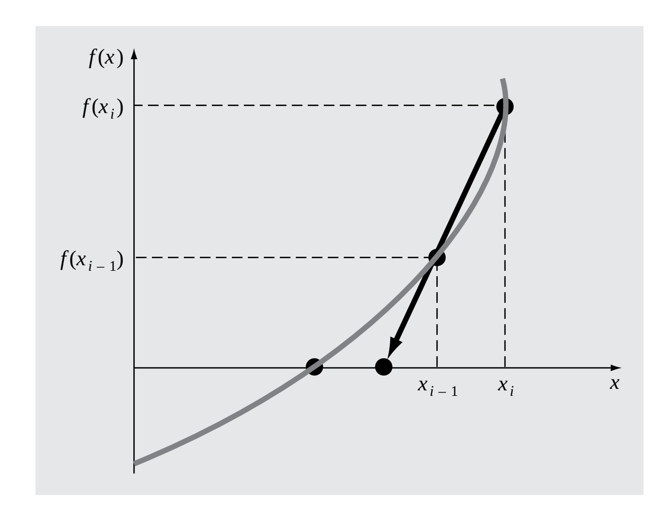


Figura 6.7
Representación gráfica del método de la secante. Esta técnica es similar a la del método de Newton-Raphson (figura 6.5) en el sentido de que una aproximación de la raíz se predice extrapolando una tangente de la función hasta el eje *x*. Sin embargo, el método de la secante usa diferencias divididas en lugar de una derivada para estimar la pendiente.

$x_0 = 1 \qquad f(x_0) = -0.63212$

$x_1 = 0.61270 \qquad f(x_1) = -0.07081$

(Note que ambas aproximaciones se encuentran del mismo lado de la raíz.)

$$x_2 = 0.61270 - \frac{-0.07081(1 - 0.61270)}{-0.63212 - (-0.07081)} = 0.56384 \qquad \varepsilon_t = 0.58\%$$

Tercera iteración:

$x_1 = 0.61270 \qquad f(x_1) = -0.07081$

$x_2 = 0.56384 \qquad f(x_2) = 0.00518$

$$x_3 = 0.56384 - \frac{0.00518(0.61270 - 0.56384)}{-0.07081 - (-0.00518)} = 0.56717 \qquad \varepsilon_t = 0.0048\%$$

6.3.1 Diferencia entre los métodos de la secante y de la falsa posición

Observe la similitud entre los métodos de la secante y de la falsa posición. Por ejemplo, las ecuaciones (5.7) y (6.7) son idénticas en todos los términos. Ambas usan dos valores iniciales para calcular una aproximación de la pendiente de la función que se utiliza para proyectar hacia el eje *x* una nueva aproximación de la raíz. Sin embargo, existe una diferencia crítica entre ambos métodos. Tal diferencia estriba en la forma en que uno de los valores iniciales se reemplaza por la nueva aproximación. Recuerde que en el método de la falsa posición, la última aproximación de la raíz reemplaza a cualquiera de los valores iniciales que dé un valor de la función con el mismo signo que $f(x_r)$. En consecuencia, las dos aproximaciones siempre encierran a la raíz. Por lo tanto, para todos los casos, el método siempre converge, pues la raíz se encuentra dentro del intervalo. En contraste, el método de la secante reemplaza los valores en secuencia estricta: con el nuevo valor x_{i+1} se reemplaza a x_i y x_i reemplaza a x_{i-1}. En consecuencia, algunas veces los dos valores están en el mismo lado de la raíz. En ciertos casos esto puede llevar a divergencias.

EJEMPLO 6.7 Comparación de la convergencia en los métodos de la secante y de la falsa posición

Planteamiento del problema Utilice los métodos de la secante y de la falsa posición para calcular la raíz de $f(x) = \ln x$. Empiece los cálculos con los valores iniciales $x_l = x_{i-1} = 0.5$ y $x_u = x_i = 5.0$.

Solución En el método de la falsa posición, con el uso de la ecuación (5.7) y los criterios del intervalo para el reemplazo de las aproximaciones, se obtienen las siguientes iteraciones:

Iteración	x_l	x_u	x_r
1	0.5	5.0	1.8546
2	0.5	1.8546	1.2163
3	0.5	1.2163	1.0585

Como se observa (figuras 6.8*a* y *c*), las aproximaciones van convergiendo a la raíz exacta que es igual a 1.

En el método de la secante, con el uso de la ecuación (6.7) y el criterio secuencial para el reemplazo de las aproximaciones, se obtiene:

Iteración	x_{i-1}	x_i	x_{i+1}
1	0.5	5.0	1.8546
2	5.0	1.8546	–0.10438

Como se muestra en la figura 6.8*d*, el método es divergente.

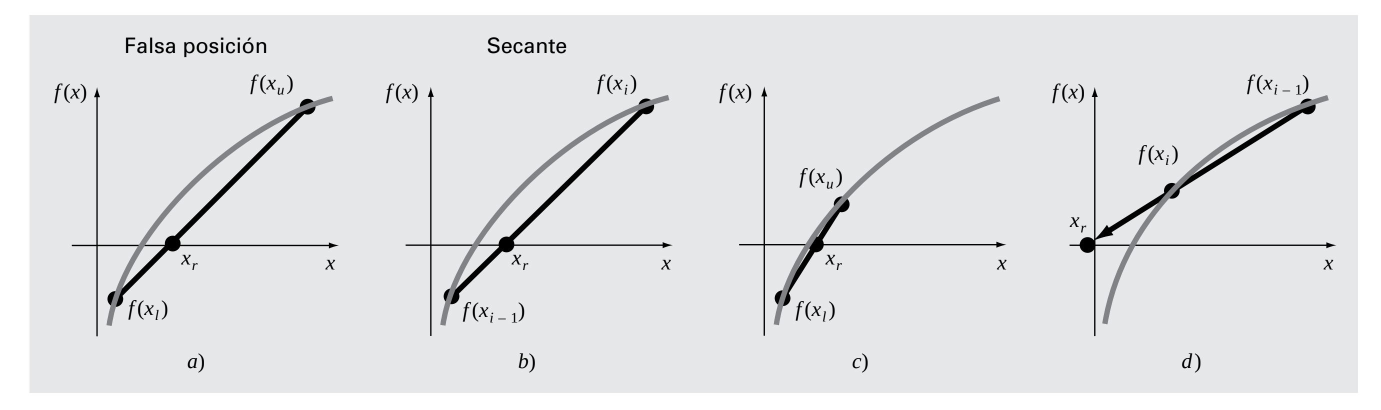


Figura 6.8
Comparación entre los métodos de la falsa posición y de la secante. Las primeras iteraciones *a*) y *b*) de ambos métodos son idénticas. No obstante, en las segundas iteraciones *c*) y *d*), los puntos usados son diferentes. En consecuencia, el método de la secante llega a diverger, como se indica en *d*).

Aunque el método de la secante sea divergente, cuando converge lo hace más rápido que el método de la falsa posición. Por ejemplo, en la figura 6.9 se muestra la superioridad del método de la secante. La inferioridad del método de la falsa posición se debe a que un extremo permanece fijo para mantener a la raíz dentro del intervalo. Esta propiedad, que es una ventaja porque previene la divergencia, tiene una desventaja en relación con la velocidad de convergencia; esto hace de la diferencia finita estimada una aproximación menos exacta que la derivada.

6.3.2 Algoritmo para el método de la secante

Como con los otros métodos abiertos, el algoritmo del método de la secante se obtiene simplemente modificando la figura 6.4, de tal forma que se puedan introducir dos valores iniciales, y usando la ecuación (6.7) se calcule la raíz. Además, las opciones sugeridas en la sección 6.2.3 para el método de Newton-Raphson, también se pueden aplicar para obtener ventajas al programa de la secante.

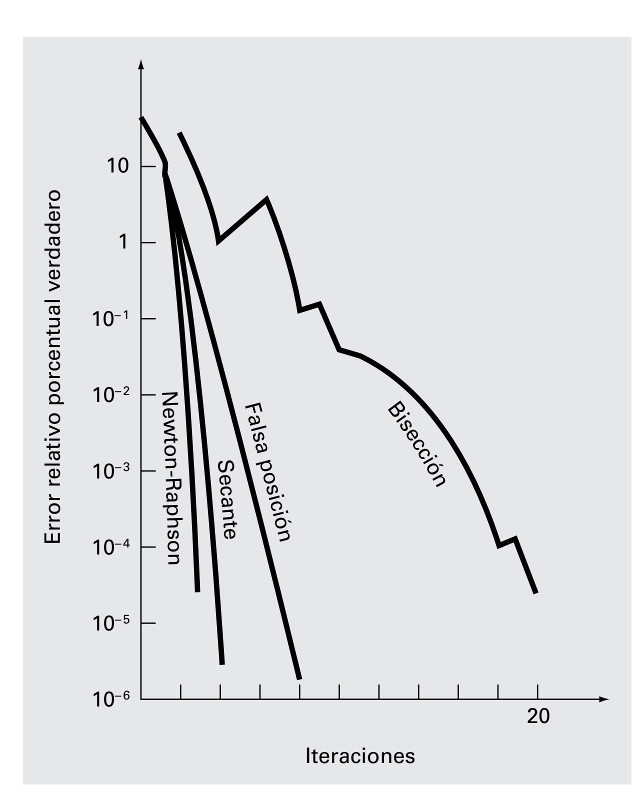


Figura 6.9
Comparación de los errores relativos porcentuales verdaderos ε_t, para los métodos que determinan las raíces de $f(x) = e^{-x} - x$.

6.3.3 Método de la secante modificado

En lugar de usar dos valores arbitrarios para aproximar la derivada, un método alternativo considera un cambio fraccionario de la variable independiente para estimar $f'(x)$,

$$f'(x_i) \cong \frac{f(x_i + \delta x_i) - f(x_i)}{\delta x_i}$$

donde δ es un pequeño cambio fraccionario. Esta aproximación se sustituye en la ecuación (6.6) que da la siguiente ecuación iterativa:

$$x_{i+1} = x_i - \frac{\delta x_i \; f(x_i)}{f(x_i + \delta x_i) - f(x_i)} \tag{6.8}$$

EJEMPLO 6.8 Método de la secante modificado

Planteamiento del problema Con el método de la secante modificado estime la raíz de $f(x) = e^{-x} - x$. Use un valor de 0.01 para δ y comience con $x_0 = 1.0$. Recuerde que la raíz verdadera es 0.56714329...

Solución

Primera iteración:

$$x_0 = 1 \qquad f(x_0) = -0.63212$$
$$x_0 + \delta x_0 = 1.01 \qquad f(x_0 + \delta x_0) = -0.64578$$

$$x_1 = 1 - \frac{0.01(-0.63212)}{-0.64578 - (-0.63212)} = 0.537263 \qquad |\varepsilon_t| = 5.3\%$$

Segunda iteración:

$$x_0 = 0.537263 \qquad f(x_0) = 0.047083$$
$$x_0 + \delta x_0 = 0.542635 \qquad f(x_0 + \delta x_0) = 0.038579$$

$$x_1 = 0.537263 - \frac{0.005373(0.047083)}{0.038579 - 0.047083} = 0.56701 \qquad |\varepsilon_t| = 0.0236\%$$

Tercera iteración:

$$x_0 = 0.56701 \qquad f(x_0) = 0.000209$$
$$x_0 + \delta x_0 = 0.572680 \qquad f(x_0 + \delta x_0) = -0.00867$$

$$x_1 = 0.56701 - \frac{0.00567(0.000209)}{-0.00867 - 0.000209} = 0.567143 \qquad |\varepsilon_t| = 2.365 \times 10^{-5}\%$$

La elección de un valor adecuado para δ no es automática. Si δ es muy pequeño, el método puede no tener éxito por el error de redondeo, causado por la cancelación por resta en el denominador de la ecuación (6.8). Si ésta es muy grande, la técnica puede llegar a ser ineficiente y hasta divergente. No obstante, si se selecciona correctamente, proporciona una adecuada alternativa en los

casos donde la evaluación de la derivada se dificulta y el desarrollo de dos valores iniciales es inconveniente.

6.4 MÉTODO DE BRENT

¿No sería agradable tener un método híbrido que combinara la confiabilidad del agrupamiento en regiones con la rapidez de los métodos abiertos? El *método de Brent de ubicación de raíces* es un ingenioso algoritmo que hace exactamente eso mediante la aplicación de un rápido método abierto siempre que sea posible, pero volviendo al confiable método de agrupamiento en regiones cuando sea necesario. Este método lo desarrolló Richard Brent (1973), con base en un algoritmo anterior de Theodorus Dekker (1969).

La técnica de agrupamiento en regiones es el confiable método de bisección (sección 5.2), mientras que se usan dos métodos abiertos distintos. El primero es el método de la secante descrito en la sección 6.3. Como se explica en seguida, el segundo es la interpolación inversa cuadrática.

6.4.1 Interpolación inversa cuadrática

La *interpolación inversa cuadrática* se basa en una idea similar a la del método de la secante. Como en la figura 6.10*a*, el método de la secante es el cálculo de una línea recta que pasa por dos valores iniciales. La intersección de esta línea recta con el eje *x* representa el nuevo estimado de la raíz. Por esta razón, algunas veces se le describe como un *método de interpolación lineal.*

Ahora suponga que se tienen tres puntos. En este caso, se puede determinar una función cuadrática de *x* que pase por los tres puntos (figura 6.10*b*). Exactamente igual que con el método de la secante lineal, la intersección de esta parábola con el eje *x* representaría el nuevo estimado de la raíz. Y, como se ilustra en la figura 6.10*b*, usar una curva en vez de una línea recta a menudo produce un mejor estimado.

Aunque esto parecería representar una gran mejora, el método tiene una falla fundamental: ¡es posible que la parábola no interseque el eje *x*! Éste sería el caso cuando la parábola resultante tenga raíces complejas. Esto se ilustra por la parábola $y = f(x)$, en la figura 6.11.

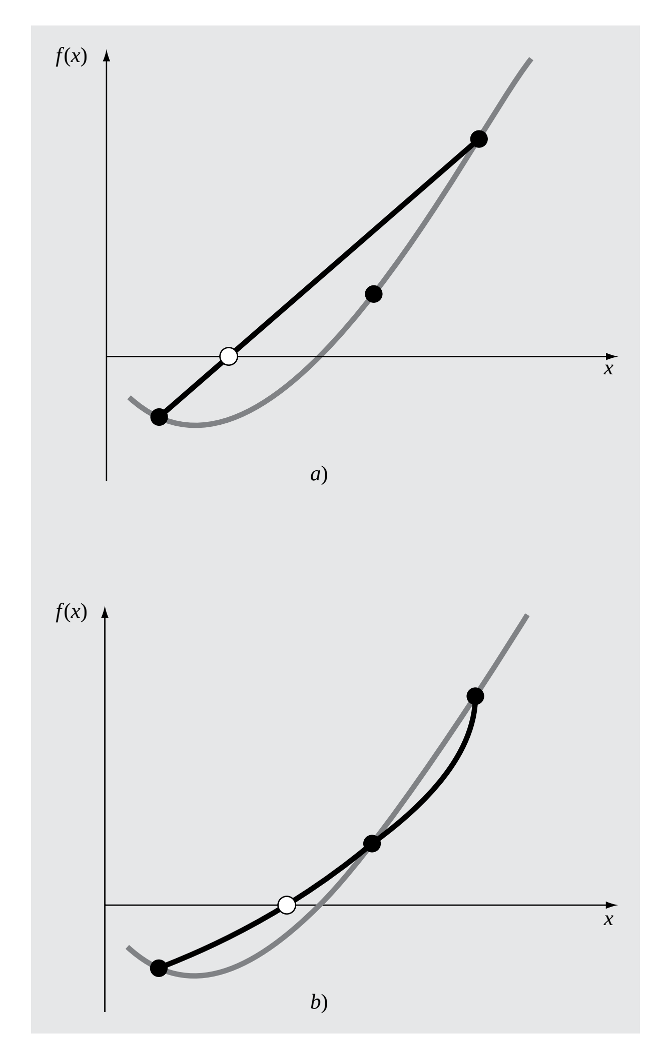


Figura 6.10
Comparación de *a*) el método de la secante y *b*) interpolación inversa cuadrática. Observe que la parábola negra que pasa por los tres puntos en *b*) se llama "inversa" porque se escribe en *y* en vez de en *x*.

La dificultad se puede corregir usando la interpolación inversa cuadrática. Es decir, en lugar de usar una parábola en *x*, se pueden ubicar los puntos en una parábola en *y*. Esto equivale a invertir los ejes y crear una parábola "ladeada" [la curva $x = f(y)$ en la figura 6.11].

Si los tres puntos se designan como (x_{i-2}, y_{i-2}), (x_{i-1}, y_{i-1}) y (x_i, y_i), se puede generar una función cuadrática de *y* que pase por los puntos como

$$g(y) = \frac{(y - y_{i-1})(y - y_i)}{(y_{i-2} - y_{i-1})(y_{i-2} - y_i)}x_{i-2} + \frac{(y - y_{i-2})(y - y_i)}{(y_{i-1} - y_{i-2})(y_{i-1} - y_i)}x_{i-1} + \frac{(y - y_{i-2})(y - y_{i-1})}{(y_i - y_{i-2})(y_i - y_{i-1})}x_i \qquad (6.9)$$

Como se verá en la sección 18.2, esta forma se llama un *polinomio de Lagrange*. La raíz, x_{i+1}, corresponde a $y = 0$, que, al sustituirse en la ecuación (6.9), da

$$x_{i+1} = \frac{y_{i-1}y_i}{(y_{i-2} - y_{i-1})(y_{i-2} - y_i)}x_{i-2} + \frac{y_{i-2}y_i}{(y_{i-1} - y_{i-2})(y_{i-1} - y_i)}x_{i-1} + \frac{y_{i-2}y_{i-1}}{(y_i - y_{i-2})(y_i - y_{i-1})}x_i \tag{6.10}$$

Como se muestra en la figura 6.11, una parábola "ladeada" como ésta siempre interseca el eje *x*.

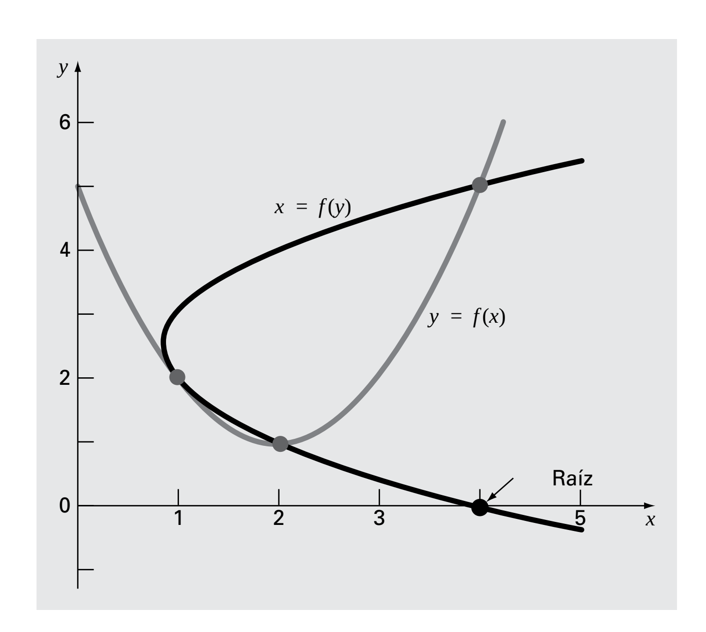


Figura 6.11
Dos parábolas que pasen por tres puntos. Parábolas escritas como función de *x*, *y* = *f*(*x*), tiene raíces complejas y por lo tanto no interseca el eje *x*. En contraste, si las variables se invierten y se desarrolla la parábola como *x* = *f*(*y*), la función sí interseca el eje *x*.

EJEMPLO 6.9 Interpolación cuadrática inversa

Planteamiento del problema Desarrolle ecuaciones cuadráticas tanto en *x* como en *y* para los puntos de datos descritos en la figura 6.11: (1, 2), (2, 1) y (4, 5). Para el primer caso, $y = f(x)$, use la fórmula cuadrática para ilustrar que las raíces son complejas. Para el segundo caso, $x = g(y)$, use la interpolación cuadrática inversa [ecuación (6.10)] para determinar el estimado de raíces.

Solución Invirtiendo las *x* y las *y*, se puede usar la ecuación (6.9) para generar una cuadrática en *x* como

$$f(x) = \frac{(x-2)(x-4)}{(1-2)(1-4)}2 + \frac{(x-1)(x-4)}{(2-1)(2-4)}1 + \frac{(x-1)(x-2)}{(4-1)(4-2)}5$$

o, agrupando términos,

$$f(x) = x^2 - 4x + 5$$

Esta ecuación se usó para generar la parábola, $y = f(x)$ en la figura 6.11. Se puede usar la fórmula cuadrática para determinar que las raíces, en este caso, son complejas,

$$x = \frac{4 \pm \sqrt{(-4)^2 - 4(1)(5)}}{2} = 2 \pm i$$

La ecuación (6.9) se puede usar para generar la cuadrática en *y* como

$$g(y) = \frac{(y-1)(y-5)}{(2-1)(2-5)}1 + \frac{(y-2)(y-5)}{(1-2)(1-5)}2 + \frac{(y-2)(y-1)}{(5-2)(5-1)}4$$

o, agrupando términos,

$$g(y) = 0.5x^2 - 2.5x + 4$$

Finalmente, se puede usar la ecuación (6.10) para determinar la raíz como

$$x_{i+1} = \frac{-1(-5)}{(2-1)(2-5)}1 + \frac{-2(-5)}{(1-2)(1-5)}2 + \frac{-2(-1)}{(5-2)(5-1)}4 = 4$$

Antes de proseguir con el algoritmo de Brent, se necesita mencionar otro caso en el que no funciona la interpolación cuadrática inversa. Si los tres valores de y no son distintos (es decir, $y_{i-2} = y_{i-1}$ o $y_{i-1} = y_i$), no existe una función cuadrática inversa. De modo que aquí es donde entra en juego el método de la secante. Si se llega a una situación en la que los valores de y no son distintos, se puede siempre volver al método menos eficiente de la secante para generar una raíz usando dos de los puntos. Si $y_{i-2} = y_{i-1}$, se usa el método de la secante con x_{i-1} y x_i. Si $y_{i-1} = y_i$, se usa x_{i-2} y x_{i-1}.

6.4.2 Algoritmo del método de Brent

La idea general respecto al *método de Brent para la ubicación de raíces* es, siempre que sea posible, usar uno de los métodos rápidos abiertos. En el caso de que esto genere un resultado inaceptable (es decir, un estimado de raíces que caiga fuera de la región seleccionada), el algoritmo vuelve al método más conservador de bisección. Aunque el método de bisección puede ser más lento, genera un estimado con garantía de caer dentro de la región. Este proceso se repite luego hasta que la raíz se ubique dentro de una tolerancia aceptable. Como podría esperarse, al principio domina usualmente la bisección, pero al acercarse a la raíz, la técnica cambia a los métodos abiertos más rápidos.

La figura 6.12 presenta el pseudocódigo para el algoritmo basado en el archivo M de MATLAB desarrollado por Cleve Moler (2005). Representa una versión simplificada de la función `fzero`, que es la función profesional para ubicación de raíces que se emplea en MATLAB. Por esta razón, la versión simplificada se llama `fzerosimp`. Observe que necesita otra función, `f`, para mantener la ecuación para la cual se está evaluando la raíz.

La función `fzerosimp` se pasa por dos valores iniciales entre los que debe estar comprendida la raíz. Después de asignar valores para la épsilon de la máquina y para una tolerancia, las tres variables que definen el intervalo de búsqueda (`a, b, c`) se inicializan y se evalúa `f` en los puntos finales.

En seguida se implementa un ciclo principal si es necesario, se reordenan los tres puntos para satisfacer las condiciones necesarias para que el algoritmo funcione eficazmente. En este punto, si se cumplen los criterios de detención, se termina el ciclo. De otra manera, una estructura de decisión elige entre los tres métodos y verifica si el resultado es aceptable. Luego, una sección final evalúa `f` en el punto nuevo y se repite el ciclo. Una vez que se cumplen los criterios de detención, el ciclo termina y se devuelve el estimado final de raíces.

Observe que la sección 7.7.2 presenta una aplicación del método de Brent en la que se ilustra cómo trabaja la función `fzero` de MATLAB. Además, se usa en el estudio de caso 8.4 para determinar el factor de fricción para flujo de aire por un tubo.

6.5 RAÍCES MÚLTIPLES

Una *raíz múltiple* corresponde a un punto donde una función es tangencial al eje x. Por ejemplo, una raíz doble resulta de

$$f(x) = (x-3)(x-1)(x-1) \tag{6.11}$$

o, multiplicando términos, $f(x) = x^3 - 5x^2 + 7x - 3$. La ecuación tiene una *raíz doble* porque un valor de x hace que dos términos de la ecuación (6.11) sean iguales a cero. Gráficamente, esto significa que la curva toca en forma tangencial al eje x en la raíz doble. Observe la figura 6.l3*a* en $x = 1$. Note que la función toca al eje pero no lo cruza en la raíz.

```
Function fzerosimp(xl, xu)
eps = 2.22044604925031E-16
tol = 0.000001
a = xl: b = xu: fa = f(a): fb = f(b)
c = a: fc = fa: d = b − c: e = d
DO
  IF fb = 0 EXIT
  IF Sgn(fa) > Sgn(fb) THEN  (Si es necesario reordene los puntos)
    a = c: fa = fc: d = b − c: e = d
  ENDIF
  IF |fa| < |fb| THEN
    c = b: b = a: a = c
    fc = fb: fb = fa: fa = fc
  ENDIF
  m = 0.5 * (a − b)    (Fin de prueba y posible salida)
  tol = 2 * eps * max(|b|, 1)
  IF |m| ≤ tol Or fb = 0. THEN
    EXIT
  ENDIF
  (Choose open methods or bisection)
  IF |e| ≥ tol And |fc| > |fb| THEN
    s = fb / fc
    IF a = c THEN                            (Método de la secante)
      p = 2 * m * s
      q = 1 − s
    ELSE                      (Interpolación cuadrática inversa)
      q = fc / fa: r = fb / fa
      p = s * (2 * m * q * (q − r) − (b − c) * (r − 1))
      q = (q − 1) * (r − 1) * (s − 1)
    ENDIF
    IF p > 0 THEN q = −q ELSE p = −p
    IF 2 * p < 3 * m * q − |tol * q| AND p < |0.5 * e * q| THEN
      e = d: d = p / q
    ELSE
      d = m: e = m
    ENDIF
  ELSE                                       (Bisección)
    d = m: e = m
  ENDIF
  c = b: fc = fb
  IF |d| > tol THEN b = b + d Else b = b − Sgn(b − a) * tol
  fb = f(b)
ENDDO
fzerosimp = b
END fzerosimp
```

Figura 6.12 Pseudocódigo para el algoritmo de Brent para búsqueda de raíces basado en un archivo M de MATLAB desarrollado por Cleve Moler (2005).

Una *raíz triple* corresponde al caso en que un valor de x hace que tres términos en una ecuación sean iguales a cero, como en

$$f(x) = (x - 3)(x - 1)(x - 1)(x - 1)$$

o, multiplicando los términos, $f(x) = x^4 - 6x^3 + 12x^2 - 10x + 3$. Advierta que la representación gráfica (figura 6.13*b*) indica otra vez que la función es tangente al eje en la raíz, pero que en este caso sí cruza el eje. En general, la multiplicidad impar de raíces cruza el eje, mientras que la multiplicidad par no lo cruza. Por ejemplo, la raíz cuádruple en la figura 6.13*c* no cruza el eje.

Las raíces múltiples ofrecen algunas dificultades a muchos de los métodos numéricos expuestos en la parte dos:

1. El hecho de que la función no cambie de signo en raíces múltiples pares impide confiarse de los métodos cerrados, que se analizan en el capítulo 5. Así, en los métodos incluidos en este texto, se está limitando a los abiertos que pueden ser divergentes.

2. Otro posible problema se relaciona con el hecho de que no sólo $f(x)$, sino también $f'(x)$ se aproxima a cero en la raíz. Tales problemas afectan los métodos de Newton-Raphson y de la secante, los cuales contienen derivadas (o su aproximación) en el denominador de sus fórmulas respectivas. Esto provocará una división entre cero cuando la solución converge muy cerca de la raíz. Una forma simple de evitar dichos problemas, que se ha demostrado teóricamente (Ralston y Rabinowitz, 1978), se basa en el hecho de que $f(x)$ siempre alcanzará un valor cero antes que $f'(x)$. Por lo tanto, si se compara $f(x)$ contra cero, dentro del programa, entonces los cálculos se pueden terminar antes de que $f'(x)$ llegue a cero.
3. Es posible demostrar que el método de Newton-Raphson y el método de la secante convergen en forma lineal, en vez de cuadrática, cuando hay raíces múltiples (Ralston y Rabinowitz, 1978). Se han propuesto algunas modificaciones para atenuar este problema. Ralston y Rabinowitz (1978) proponen que se realice un pequeño cambio en la formulación para que se regrese a la convergencia cuadrática, como en

$$x_{i+1} = x_i - m\frac{f(x_i)}{f'(x_i)} \tag{6.12}$$

donde m es la multiplicidad de la raíz (es decir, $m = 2$ para una raíz doble, $m = 3$ para una raíz triple, etc.). Se trata de una alternativa poco satisfactoria, porque depende del conocimiento de la multiplicidad de la raíz.

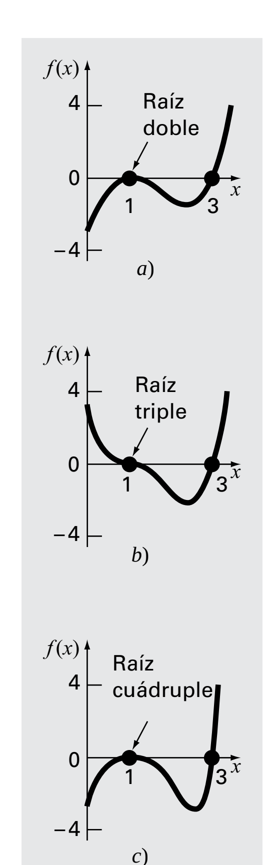


Figura 6.13
Ejemplos de raíces múltiples que son tangenciales al eje x. Observe que la función no cruza el eje en los casos de raíces múltiples pares *a*) y *c*), mientras que con multiplicidad impar sí lo hace en *b*).

Otra alternativa, también sugerida por Ralston y Rabinowitz (1978), consiste en definir una nueva función $u(x)$, que es el cociente de la función original entre su derivada:

$$u(x) = \frac{f(x)}{f'(x)} \tag{6.13}$$

Se puede demostrar que esta función tiene raíces en las mismas posiciones que la función original. Por lo tanto, la ecuación (6.13) se sustituye en la ecuación (6.6) para desarrollar una forma alternativa del método de Newton-Raphson:

$$x_{i+1} = x_i - \frac{u(x_i)}{u'(x_i)} \tag{6.14}$$

La ecuación (6.13) se puede diferenciar para obtener

$$u'(x) = \frac{f'(x)f'(x) - f(x)f''(x)}{[f'(x)]^2} \tag{6.15}$$

Se sustituyen las ecuaciones (6.13) y (6.15) en la ecuación (6.14) y se simplifica el resultado:

$$x_{i+1} = x_i - \frac{f(x_i)f'(x_i)}{[f'(x)]^2 - f(x_i)f''(x_i)} \tag{6.16}$$



EJEMPLO 6.10 Método de Newton-Raphson modificado para el cálculo de raíces múltiples

Planteamiento del problema Con los dos métodos, el estándar y el modificado, de Newton-Raphson evalúe la raíz múltiple de la ecuación (6.11), use un valor inicial de $x_0 = 0$.

Solución La primera derivada de la ecuación (6.11) es $f'(x) = 3x^2 - 10x + 7$, y por lo tanto, el método de Newton-Raphson estándar para este problema es [ecuación (6.6)]

$$x_{i+1} = x_i - \frac{x_i^3 - 5x_i^2 + 7x_i - 3}{3x_i^2 - 10x_i + 7}$$

que se resuelve iterativamente para obtener

i	x_i	ε_t (%)
0	0	100
1	0.4285714	57
2	0.6857143	31
3	0.8328654	17
4	0.9133290	8.7
5	0.9557833	4.4
6	0.9776551	2.2

Como ya se había anticipado, el método converge en forma lineal hacia el valor verdadero 1.0.

Para el caso del método modificado, la segunda derivada es $f''(x) = 6x - 10$, y en consecuencia la ecuación iterativa será [ecuación (6.16)]

$$x_{i+1} = x_i - \frac{(x_i^3 - 5x_i^2 + 7x_i - 3)(3x_i^2 - 10x_i + 7)}{(3x_i^2 - 10x_i + 7)^2 - (x_i^3 - 5x_i^2 + 7x_i - 3)(6x_i - 10)}$$

que se resuelve para obtener

i	x_i	ε_t (%)
0	0	100
1	1.105263	11
2	1.003082	0.31
3	1.000002	0.00024

De esta manera, la fórmula modificada converge en forma cuadrática. Se pueden usar ambos métodos para buscar la raíz simple en $x = 3$. Con un valor inicial $x_0 = 4$ se obtienen los siguientes resultados:

i	Estándar	ε_t (%)	Modificado	ε_t (%)
0	4	33	4	33
1	3.4	13	2.636364	12
2	3.1	3.3	2.820225	6.0
3	3.008696	0.29	2.961728	1.3
4	3.000075	0.0025	2.998479	0.051
5	3.000000	2×10^{-7}	2.999998	7.7×10^{-5}

De esta forma, deberá notar que ambos métodos convergen con rapidez, aunque el método estándar es el más eficiente.

En el ejemplo anterior se ilustran los factores de mayor importancia involucrados al elegir el método de Newton-Raphson modificado. Aunque es preferible para raíces múltiples, es menos eficiente y requiere más trabajo computacional que el método estándar para raíces simples.

Se debe notar que hay manera de desarrollar una versión modificada del método de la secante para raíces múltiples, sustituyendo la ecuación (6.13) en la ecuación (6.7). La fórmula resultante es (Ralston y Rabinowitz, 1978)

$$x_{i+1} = x_i - \frac{u(x_i)(x_{i-1} - x_i)}{u(x_{i-1}) - u(x_i)}$$

6.6 SISTEMAS DE ECUACIONES NO LINEALES

Hasta aquí nos hemos ocupado de determinar las raíces de una sola ecuación no lineal. Un problema relacionado con éste consiste en obtener las raíces de un conjunto de ecuaciones simultáneas,

$$\begin{aligned} f_1(x_1, x_2, ..., x_n) &= 0 \\ f_2(x_1, x_2, ..., x_n) &= 0 \\ &\vdots \\ f_n(x_1, x_2, ..., x_n) &= 0 \end{aligned} \tag{6.17}$$

La solución de este sistema consta de un conjunto de valores x que simultáneamente hacen que todas las ecuaciones sean iguales a cero.

En la parte tres, presentaremos los métodos, para el caso en que las ecuaciones simultáneas son lineales, es decir, que se puedan expresar en la forma general

$$f(x) = a_1x_1 + a_2x_2 + \cdots + a_nx_n - b = 0 \tag{6.18}$$

donde la b y las a son constantes. A las ecuaciones algebraicas y trascendentes que no se pueden expresar de esta forma se les llama *ecuaciones no lineales*. Por ejemplo,

$$x^2 + xy = 10$$

y

$$y + 3xy^2 = 57$$

son dos ecuaciones simultáneas no lineales con dos incógnitas, x y y, las cuales se expresan en la forma de la ecuación (6.17) como

$$u(x, y) = x^2 + xy - 10 = 0 \tag{6.19a}$$

$$v(x, y) = y + 3xy^2 - 57 = 0 \tag{6.19b}$$

Así, la solución serían los valores de x y de y que hacen a las funciones $u(x, y)$ y $v(x, y)$ iguales a cero. La mayoría de los métodos para determinar tales soluciones son extensiones de los métodos abiertos para resolver ecuaciones simples. En esta sección presentaremos dos de ellos: iteración de punto fijo y Newton-Raphson.

6.6.1 Iteración de punto fijo

El método de iteración de punto fijo (sección 6.1) puede modificarse para resolver dos ecuaciones simultáneas no lineales. Este método se ilustra en el siguiente ejemplo.

EJEMPLO 6.11 Iteración de punto fijo para un sistema no lineal

Planteamiento del problema Con el método de iteración de punto fijo determine las raíces de la ecuación (6.19). Observe que un par correcto de raíces es $x = 2$ y $y = 3$. Inicie el cálculo con el valor inicial $x = 1.5$ y $y = 3.5$.

Solución De la ecuación (6.19a) se puede despejar

$$x_{i+1} = \frac{10 - x_i^2}{y_i} \tag{E6.11.1}$$

y de la ecuación (6.19b) se puede despejar

$$y_{i+1} = 57 - 3x_iy_i^2 \tag{E6.11.2}$$

Observe que dejaremos los subíndices en el resto del ejemplo.

Con base en los valores iniciales, la ecuación (E6.11.1) se utiliza para determinar un nuevo valor de x:

$$x = \frac{10 - (1.5)^2}{3.5} = 2.21429$$

Este resultado y el valor inicial de $y = 3.5$ se sustituye en la ecuación (E6.11.2) para determinar un nuevo valor de y:

$$y = 57 - 3(2.21429)(3.5)^2 = -24.37516$$

Así, parece que el método diverge. Este comportamiento es aún más pronunciado en la segunda iteración:

$$x = \frac{10 - (2.21429)^2}{-24.37516} = -0.20910$$

$$y = 57 - 3(-0.20910)(-24.37516)^2 = 429.709$$

En efecto, la aproximación se está descomponiendo.

Ahora repita el cálculo, pero con la ecuación original puesta en una forma diferente. Por ejemplo, un despeje alternativo de la ecuación (6.19*a*) es

$$x = \sqrt{10 - xy}$$

y de la ecuación (6.19*b*) es

$$y = \sqrt{\frac{57 - y}{3x}}$$

Ahora los resultados son más satisfactorios:

$$x = \sqrt{10 - 1.5(3.5)} = 2.17945$$

$$y = \sqrt{\frac{57 - 3.5}{3(2.17945)}} = 2.86051$$

$$x = \sqrt{10 - 2.17945(2.86051)} = 1.94053$$

$$y = \sqrt{\frac{57 - 2.86051}{3(1.94053)}} = 3.04955$$

Así, la aproximación converge hacia la solución correcta $x = 2$ y $y = 3$.

El ejemplo anterior ilustra la más seria desventaja de la iteración simple de punto fijo, ésta es que, la convergencia depende de la manera en que se formula la ecuación. Además, aun cuando la convergencia es posible, la divergencia puede ocurrir si los valores iniciales no son suficientemente cercanos a la solución verdadera. Usando un razonamiento similar al del cuadro 6.1, se demuestra que las condiciones suficientes para la convergencia en el caso de dos ecuaciones son

$$\left|\frac{\partial u}{\partial x}\right| + \left|\frac{\partial u}{\partial y}\right| < 1$$

y

$$\left|\frac{\partial v}{\partial x}\right| + \left|\frac{\partial v}{\partial y}\right| < 1$$

Estos criterios son tan restringidos que el método de punto fijo tiene una utilidad limitada para resolver sistemas no lineales. Sin embargo, como se describirá más adelante en el libro, será muy útil para resolver sistemas de ecuaciones lineales.

6.6.2 Newton-Raphson

Recuerde que el método de Newton-Raphson se utilizó empleando la derivada (al evaluar, es la pendiente de la recta tangente) de una función, para calcular su intersección con el eje de la variable independiente; esto es, la raíz (figura 6.5). Dicho cálculo se basó en la expansión de la serie de Taylor de primer orden (recuerde el cuadro 6.2),

$$f(x_{i+1}) = f(x_i) + (x_{i+1} - x_i)\, f'(x_i) \tag{6.20}$$

donde x_i es el valor inicial de la raíz y x_{i+1} es el valor en el cual la recta tangente interseca el eje x. En esta intersección, $f(x_{i+1})$ es, por definición, igual a cero y la ecuación (6.20) se reordena para tener

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \tag{6.21}$$

que es la forma del método de Newton-Raphson para una sola ecuación.

La forma para múltiples ecuaciones se obtiene en forma idéntica. Sin embargo, se debe usar una serie de Taylor de múltiples variables para tomar en cuenta el hecho de que más de una variable independiente contribuye a la determinación de la raíz. En el caso de dos variables, una serie de Taylor de primer orden se escribe [recuerde la ecuación (4.26)] para cada ecuación no lineal como

$$u_{i+1} = u_i + (x_{i+1} - x_i)\frac{\partial u_i}{\partial x} + (y_{i+1} - y_i)\frac{\partial u_i}{\partial y} \tag{6.22a}$$

y

$$v_{i+1} = v_i + (x_{i+1} - x_i)\frac{\partial v_i}{\partial x} + (y_{i+1} - y_i)\frac{\partial v_i}{\partial y} \tag{6.22b}$$

De la misma manera como en la versión para una sola ecuación, la raíz aproximada corresponde a los valores de x y y, donde u_{i+1} y v_{i+1} son iguales a cero. En tal situación, se reordena la ecuación (6.22) como:

$$\frac{\partial u_i}{\partial x} x_{i+1} + \frac{\partial u_i}{\partial y} y_{i+1} = -u_i + x_i\frac{\partial u_i}{\partial x} + y_i\frac{\partial u_i}{\partial y} \tag{6.23a}$$

$$\frac{\partial v_i}{\partial x} x_{i+1} + \frac{\partial v_i}{\partial y} y_{i+1} = -v_i + x_i\frac{\partial v_i}{\partial x} + y_i\frac{\partial v_i}{\partial y} \tag{6.23b}$$

Debido a que se conocen todos los valores con subíndice i (corresponden al último valor estimado), las únicas incógnitas son x_{i+1} y y_{i+1}. Entonces, la ecuación (6.23) es un conjunto de dos ecuaciones lineales con dos incógnitas [compare con la ecuación (6.18)]. En consecuencia, se pueden usar manipulaciones algebraicas (por ejemplo, la regla de Cramer) para resolverlo:

$$x_{i+1} = x_i - \frac{u_i\dfrac{\partial v_i}{\partial y} - v_i\dfrac{\partial u_i}{\partial y}}{\dfrac{\partial u_i}{\partial x}\dfrac{\partial v_i}{\partial y} - \dfrac{\partial u_i}{\partial y}\dfrac{\partial v_i}{\partial x}} \tag{6.24a}$$

$$y_{i+1} = y_i - \frac{v_i \frac{\partial u_i}{\partial x} - u_i \frac{\partial v_i}{\partial x}}{\frac{\partial u_i}{\partial x}\frac{\partial v_i}{\partial y} - \frac{\partial u_i}{\partial y}\frac{\partial v_i}{\partial x}} \tag{6.24b}$$

El denominador de cada una de esas ecuaciones se conoce formalmente como el determinante *jacobiano* del sistema.

La ecuación (6.24) es la versión para dos ecuaciones del método de Newton-Raphson. Como en el siguiente ejemplo, se puede emplear en forma iterativa para determinar las raíces de dos ecuaciones simultáneas.

EJEMPLO 6.12 Newton-Raphson para un sistema no lineal

Planteamiento del problema Con el método de Newton-Raphson para múltiples ecuaciones determine las raíces de la ecuación (6.19). Observe que un par correcto de raíces es $x = 2$ y $y = 3$. Use como valores iniciales $x = 1.5$ y $y = 3.5$.

Solución Primero calcule las derivadas parciales y evalúelas con los valores iniciales de x y y:

$$\frac{\partial u_0}{\partial x} = 2x + y = 2(1.5) + 3.5 = 6.5 \qquad \frac{\partial u_0}{\partial y} = x = 1.5$$

$$\frac{\partial v_0}{\partial x} = 3y^2 = 3(3.5)^2 = 36.75 \qquad \frac{\partial v_0}{\partial y} = 1 + 6xy = 1 + 6(1.5)(3.5) = 32.5$$

Así, el determinante jacobiano para la primera iteración es

$$6.5(32.5) - 1.5(36.75) = 156.125$$

Los valores de las funciones se evalúan con los valores iniciales como

$$u_0 = (1.5)^2 + 1.5(3.5) - 10 = -2.5$$

$$v_0 = 3.5 + 3(1.5)(3.5)^2 - 57 = 1.625$$

Estos valores se sustituyen en la ecuación (6.24):

$$x = 1.5 - \frac{-2.5(32.5) - 1.625(1.5)}{156.125} = 2.03603$$

$$y = 3.5 - \frac{1.625(6.5) - (-2.5)(36.75)}{156.125} = 2.84388$$

Así, los resultados están convergiendo a los valores verdaderos $x = 2$ y $y = 3$. Los cálculos se repiten hasta que se obtenga una precisión aceptable.

Como con el método de iteración de punto fijo, la aproximación de Newton-Raphson puede diverger si los valores iniciales no están lo suficientemente cercanos a la raíz verdadera. Mientras que para el caso de una sola ecuación los métodos gráficos son útiles para obtener un buen valor inicial, ningún procedimiento tan simple está disponible para el caso de múltiples ecuaciones. Aunque existen algunos métodos avanzados para obtener una primera aproximación aceptable, los valores iniciales a menudo deben obtenerse mediante prueba y error, con el conocimiento del sistema físico que se está modelando.

El método de Newton-Raphson para dos ecuaciones puede generalizarse para resolver n ecuaciones simultáneas. Debido a que el camino más eficiente para esto implica el álgebra matricial y la solución de ecuaciones lineales simultáneas, se pospondrá su estudio para la parte tres.

PROBLEMAS

6.1 Utilice la iteración simple de punto fijo para localizar la raíz de

$$f(x) = \text{sen}\left(\sqrt{x}\right) - x$$

Haga una elección inicial de $x_0 = 0.5$ e itere hasta que $\varepsilon_a \le 0.01\%$. Compruebe que el proceso converge en forma lineal según se describió en el cuadro 6.1.

6.2 Determine la raíz real más grande de

$$f(x) = 2x^3 - 11.7x^2 + 17.7x - 5$$

a) En forma gráfica.
b) Con el método de iteración simple de punto fijo (tres iteraciones, $x_0 = 3$). *Nota:* Asegúrese de haber desarrollado una solución que converja a la raíz.
c) Con el método de Newton-Raphson (tres iteraciones, $x_0 = 3$).
d) Con el método de la secante (tres iteraciones $x_{-1} = 3$, $x_0 = 4$).
e) Con el método de la secante modificado (tres iteraciones, $x_0 = 3$, $\delta = 0.01$). Calcule el porcentaje aproximado de errores relativos para sus soluciones.

6.3 Utilice los métodos de *a*) iteración de punto fijo y *b*) Newton-Raphson, para determinar una raíz de $f(x) = -0.9x^2 + 1.7x + 2.5$ con el uso de $x_0 = 5$. Haga el cálculo hasta que ε_a sea menor que $\varepsilon_s = 0.01\%$. Asimismo, realice una comprobación del error de su respuesta final.

6.4 Determine las raíces reales de $f(x) = -1 + 5.5x - 4x^2 + 0.5x^3$: *a*) en forma gráfica y *b*) con el método de Newton-Raphson dentro de $\varepsilon_s = 0.01\%$.

6.5 Emplee el método de Newton-Raphson para determinar una raíz real de $f(x) = -1 + 5.5x - 4x^2 + 0.5x^3$ con el uso de valores iniciales de *a*) 4.52 y *b*) 4.54. Estudie y use métodos gráficos y analíticos para explicar cualquier peculiaridad en sus resultados.

6.6 Determine la raíz real más pequeña de $f(x) = -12 - 21x + 18x^2 - 2.4x^3$: *a*) en forma gráfica y *b*) con el empleo del método de la secante para un valor de ε_s que corresponda a tres cifras significativas.

6.7 Localice la primera raíz positiva de

$$f(x) = \text{sen}\, x + \cos(1 + x^2) - 1$$

donde *x* está en radianes. Para localizar la raíz, use cuatro iteraciones del método de la secante con valores iniciales de *a*) $x_{i-1} = 1.0$ y $x_i = 3.0$; *b*) $x_{i-1} = 1.5$ y $x_i = 2.5$, *c*) $x_{i-1} = 1.5$ y $x_i = 2.25$, *d*) use el método gráfico para explicar su resultado.

6.8 Determine la raíz real de $x^{3.5} = 80$, con el método de la secante modificado dentro de $\varepsilon_s = 0.1\%$, con el uso de una elección inicial de $x_0 = 3.5$ y $\delta = 0.01$.

6.9 Determine la raíz real más grande de $f(x) = x^3 - 6x^2 + 11x - 6.1$.
a) En forma gráfica.
b) Con el uso del método de Newton-Raphson (tres iteraciones, $x_i = 3.5$).
c) Con el método de la secante (tres iteraciones, $x_{i+1} = 2.5$ y $x_i = 3.5$).
d) Por medio del método de la secante modificado (tres iteraciones, $x_i = 3.5$, $\delta = 0.01$).

6.10 Determine la menor raíz positiva de $f(x) = 7\,\text{sen}(x)e^{-x} - 1$:
a) En forma gráfica.
b) Con el uso del método de Newton-Raphson (tres iteraciones, $x_i = 0.3$).
c) Con el método de la secante (cinco iteraciones, $x_{i-1} = 0.5$ y $x_i = 0.4$).
d) Por medio del método de la secante modificado (tres iteraciones $x_i = 0.3$, $\delta = 0.01$).

6.11 Utilice el método de Newton-Raphson para encontrar la raíz de

$$f(x) = e^{-0.5x}(4 - x) - 2$$

Utilice conjeturas iniciales de *a*) 2, *b*) 6 y *c*) 8. Explique sus resultados.

6.12 Dado que

$$f(x) = -2x^6 - 1.5x^4 + 10x + 2$$

use una técnica de ubicación de raíces para determinar el máximo de esta función. Realice iteraciones hasta que el error relativo aproximado caiga por debajo de 5%. Si emplea un método de agrupamiento en regiones, use valores iniciales de $x_l = 0$ y $x_u = 1$. Si usted utiliza el método Newton-Raphson o el método modificado de la secante, use un valor inicial de $x_i = 1$. Si se decide por el método de la secante, use valores iniciales de $x_{i-1} = 0$ y $x_i = 1$. Suponiendo que no importa la convergencia, elija la técnica que sea más adecuada para este problema. Justifique su elección.

6.13 Usted debe determinar la raíz de la siguiente función fácilmente diferenciable,

$$e^{0.5x} = 5 - 5x$$

Elija la mejor técnica numérica, justifique su elección y luego use esa técnica para determinar la raíz. Observe que se sabe que, para valores iniciales positivos, todas las técnicas, salvo la iteración de punto fijo, finalmente convergen. Realice iteraciones hasta que el error relativo aproximado caiga por debajo de 2%. Si usted emplea un método cerrado, use valores iniciales de $x_l = 0$ y $x_u = 2$. Si escoge el método de Newton-Raphson o el modificado de secante, use un valor inicial de $x_i = 0.7$. Si se decide por el método de secante, use valores iniciales de $x_{i-1} = 0$ y $x_i = 2$.

6.14 Use *a*) el método Newton-Raphson y *b*) el método de secante modificada ($\delta = 0.05$) para determinar una raíz de $f(x) = x^5 - 16.05x^4 + 88.75x^3 - 192.0375x^2 + 116.35x + 31.6875$ usando un valor inicial de $x = 0.5825$ y $\varepsilon_s = 0.01\%$. Explique sus resultados.

6.15 El antiguo método de *dividir y promediar*, para obtener una apoximación de la raíz cuadrada de cualquier número positivo, *a*, se formula del modo siguiente:

$$x = \frac{x + a/x}{2}$$

Demuestre que éste es equivalente al algoritmo de Newton-Raphson.

6.16 *a*) Aplique el método de Newton-Raphson a la función $f(x) = \tanh(x^2 - 9)$ para evaluar su raíz real conocida en $x = 3$. Use un valor inicial de $x_0 = 3.2$ y haga un mínimo de cuatro iteraciones. *b*) ¿Converge el método a su raíz real? Bosqueje la gráfica con los resultados para cada iteración que obtenga.

6.17 El polinomio $f(x) = 0.0074x^4 - 0.284x^3 + 3.355x^2 - 12.183x + 5$ tiene una raíz real entre 15 y 20. Aplique el método de Newton-Raphson a dicha función con valor inicial $x_0 = 16.15$. Explique sus resultados.

6.18 Emplee el método de la secante con la función del círculo $(x + 1)^2 + (y - 2)^2 = 16$, a fin de encontrar una raíz real positiva. Haga que el valor inicial sea $x_i = 3$ y $x_{i-1} = 0.5$. Aproxímese a la solución del primer y cuarto cuadrantes. Cuando resuelva para $f(x)$ en el cuarto

cuadrante, asegúrese de tomar el valor negativo de la raíz cuadrada. ¿Por qué diverge la solución?

6.19 Suponga el lector que está diseñando un tanque esférico (véase la figura P6.19) de almacenamiento de agua para un poblado pequeño de un país en desarrollo. El volumen del líquido que puede contener se calcula con

$$V = \pi h^2 \frac{[3R - h]}{3}$$

donde V = volumen [m^3], h = profundidad del agua en el tanque [m] y R = radio del tanque [m].

Si R = 3 m, ¿a qué profundidad debe llenarse el tanque de modo que contenga 30 m^3? Haga tres iteraciones del método de Newton-Raphson para determinar la respuesta. Encuentre el error relativo aproximado después de cada iteración. Observe que el valor inicial de R convergerá siempre.

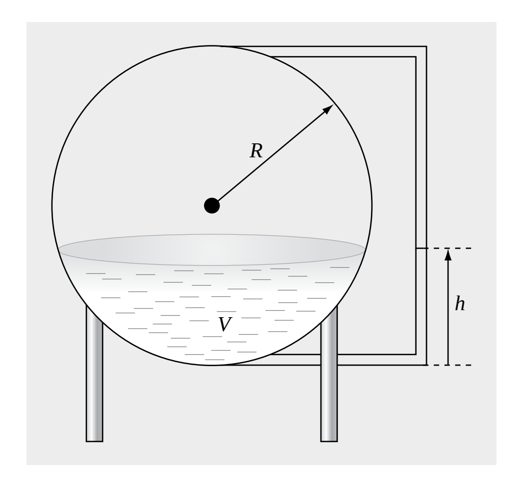


Figura P6.19

6.20 La ecuación de Manning se puede escribir para un flujo en un canal abierto rectangular como

$$Q = \frac{\sqrt{S}(BH)^{5/3}}{n(B + 2H)^{2/3}}$$

donde Q = flujo [m^3/s], S = pendiente [m/m], H = profundidad [m] y n = coeficiente de rugosidad de Manning. Desarrolle un esquema de iteración para despejar H de esta ecuación dado Q = 5, S = 0.0002, B = 20 y n = 0.03. Pruebe que su esquema converge para todos los valores iniciales mayores que cero o iguales a cero.

6.21 La función $x^3 - 2x^2 - 4x + 8$ tiene una raíz doble en x = 2. Emplee *a*) el método estándar de Newton-Raphson [ecuación (6.6)], *b*) el método de Newton-Raphson modificado [ecuación (6.12)] y *c*) el método de Newton-Raphson modificado [ecuación (6.16)] para resolver para la raíz en x = 2. Compare y analice la velocidad de convergencia con un valor inicial x_0 = 1.2.

6.22 Determine las raíces de las siguientes ecuaciones no lineales simultáneas, por medio de los métodos de *a*) iteración de punto fijo y *b*) Newton-Raphson:

$$y = -x^2 + x + 0.75$$
$$y + 5xy = x^2$$

Utilice valores iniciales de $x = y = 1.2$, y analice los resultados.

6.23 Encuentre las raíces de las ecuaciones simultáneas que siguen:

$$(x - 4)^2 + (y - 4)^2 = 5$$
$$x^2 + y^2 = 16$$

Use un enfoque gráfico para obtener los valores iniciales. Encuentre estimaciones refinadas con el método de Newton-Raphson para dos ecuaciones, que se describe en la sección 6.6.2.

6.24 Repita el problema 6.23, pero ahora determinando la raíz positiva de

$$y = x^2 + 1$$
$$y = 2 \cos x$$

6.25 El balance de masa de un contaminante en un lago bien mezclado se expresa así:

$$V \frac{dc}{dt} = W - Qc - kV\sqrt{c}$$

Dados los valores de parámetros $V = 1 \times 10^6$ m^3, $Q = 1 \times 10^5$ m^3/año, $W = 1 \times 10^6$ g/año y $k = 0.25$ m$^{0.5}$/g$^{0.5}$/año, use el método de la secante modificado para resolver para la concentración de estado estable. Emplee un valor inicial c = 4 g/m^3 y δ = 0.5. Realice tres iteraciones y determine el error relativo porcentual después de la tercera iteración.

6.26 Para el problema 6.25, la raíz puede localizarse con iteración de punto fijo como

$$c = \left(\frac{W - Qc}{kV}\right)^2$$

o bien como

$$c = \frac{W - kV\sqrt{c}}{Q}$$

De las que sólo una convergerá para valores iniciales de $2 < c < 6$. Seleccione la correcta y demuestre por qué siempre lo será.

6.27 Desarrolle un programa amigable para el usuario para el método de Newton-Raphson, con base en la figura 6.4 y la sección 6.2.3. Pruébelo por medio de repetir el cálculo del ejemplo 6.3.

6.28 Desarrolle un programa amigable para el usuario para el método de la secante, con base en la figura 6.4 y la sección 6.3.2. Pruébelo con la repetición de los cálculos del ejemplo 6.6.

6.29 Haga un programa amigable para el usuario para el método de la secante modificado, con base en la figura 6.4 y la sección 6.3.2. Pruébelo con la repetición del cálculo del ejemplo 6.8.

6.30 Desarrolle un programa fácil para el usuario del método de Brent para la ubicación de raíces, basado en la figura 6.12. Pruébelo resolviendo el problema 6.6.

6.31 Desarrolle un programa amigable para el usuario para el método de Newton-Raphson para dos ecuaciones, con base en la sección 6.6.2. Pruébelo con la solución del ejemplo 6.12.

6.32 Use el programa que desarrolló en el problema 6.31 para resolver los problemas 6.22 y 6.23, con una tolerancia de ε_s = 0.01%.

CAPÍTULO 7

Raíces de polinomios

En este capítulo estudiaremos los métodos para encontrar las raíces de ecuaciones polinomiales de la forma general

$$f_n(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \tag{7.1}$$

donde n es el grado del polinomio y las a son los coeficientes constantes. Aunque los coeficientes pueden ser números complejos, se limitará la discusión a los casos en que son reales. Para tales casos, las raíces del polinomio pueden ser reales y/o complejas.

Las raíces de los polinomios cumplen estas reglas:

1. En una ecuación de grado n, hay n raíces reales o complejas. Se debe notar que esas raíces no necesariamente son distintas.
2. Si n es impar, hay al menos una raíz real.
3. Si existen raíces complejas, éstas se encuentran por pares conjugados (es decir, $\lambda + \mu i$ y $\lambda - \mu i$), donde $i = \sqrt{-1}$.

Antes de describir las técnicas para localizar las raíces de polinomios, se proporcionarán algunos antecedentes. La primera sección da una motivación para estudiar dichas técnicas; la segunda trata de algunas manipulaciones computacionales fundamentales con polinomios.

7.1 POLINOMIOS EN LA CIENCIA Y EN LA INGENIERÍA

Los polinomios tienen muchas aplicaciones en la ciencia y en la ingeniería. Por ejemplo, se usan mucho en el ajuste de curvas. Aunque se considera que una de las aplicaciones más interesantes y potentes es la caracterización de sistemas dinámicos y, en particular, de sistemas lineales. Algunos ejemplos son los dispositivos mecánicos, las estructuras y los circuitos eléctricos. Se analizarán ejemplos específicos en el resto del texto. Éstos, en particular, se enfocarán a varias aplicaciones en la ingeniería.

Por ahora se mantendrá una discusión simple y general estudiando un sistema físico de segundo grado modelado con la siguiente *ecuación diferencial ordinaria* (EDO) lineal:

$$a_2 \frac{d^2y}{dt^2} + a_1 \frac{dy}{dt} + a_0 y = F(t) \tag{7.2}$$

donde y y t son las variables dependiente e independiente, respectivamente, las a son coeficientes constantes y $F(t)$ es la función de fuerza (término no homogéneo).

Además, se debe observar que la ecuación (7.2) puede expresarse en forma alternativa transformándola en un par de EDO de primer grado, mediante la definición de una nueva variable z,

$$z = \frac{dy}{dt} \tag{7.3}$$

La ecuación (7.3) se sustituye con su derivada en la ecuación (7.2) para eliminar el término de la segunda derivada. Esto reduce el problema a resolver

$$\frac{dz}{dt} = \frac{F(t) - a_1 z - a_0 y}{a_2} \tag{7.4}$$

$$\frac{dy}{dt} = z \tag{7.5}$$

En forma similar, una EDO lineal de grado *n*-ésimo siempre puede transformarse en un sistema de *n* EDO de primer grado.

Ahora veamos la solución. La función de fuerza representa el efecto del mundo exterior sobre el sistema. La *solución general* de la ecuación homogénea trata el caso donde la función de fuerza es igual a cero,

$$a_2 \frac{d^2y}{dt^2} + a_1 \frac{dy}{dt} + a_0 y = 0 \tag{7.6}$$

Entonces, como su nombre lo indica, la *solución general* describe algo muy general acerca del sistema que está simulando; es decir, cómo responde el sistema en ausencia de un estímulo externo.

Ahora bien, como la solución general de todos los sistemas lineales no forzados es de la forma $y = e^{rt}$. Si esta función se deriva y se sustituye en la ecuación (7.6), el resultado es

$$a_2r^2e^{rt} + a_1re^{rt} + a_0e^{rt} = 0$$

cancelando los términos exponenciales,

$$a_2r^2 + a_1r + a_0 = 0 \tag{7.7}$$

Observe que el resultado es un polinomio llamado *ecuación característica*. Las raíces de este polinomio son los valores de r que satisfacen la ecuación (7.7). Las r se conocen como los valores característicos, o *eigenvalores*, del sistema.

Se tiene aquí la relación entre las raíces de polinomios con la ciencia y la ingeniería. Los eigenvalores nos dicen algo fundamental acerca del sistema que se está modelando, así encontrar los eigenvalores implica encontrar las raíces de los polinomios. Y mientras encontrar las raíces de una ecuación de segundo grado es fácil con la fórmula cuadrática, encontrar las raíces de sistemas de grado superior (y, por lo tanto, de un polinomio de grado superior) es arduo desde el punto de vista analítico. Entonces, se requiere usar métodos numéricos del tipo descrito en este capítulo.

Antes de proceder con dichos métodos, investigaremos más profundamente qué valores específicos de los eigenvalores están implicados en el comportamiento de sistemas físicos. Primero se evaluarán las raíces de la ecuación (7.7) con la fórmula cuadrática

$$\begin{matrix} r_1 \\ r_2 \end{matrix} = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_2a_0}}{a_0}$$

Se obtienen dos raíces. Si el *discriminante* ($a_1^2 - 4a_2a_0$) es positivo, las raíces son reales y la solución general se representa como

$$y = c_1e^{r_1t} + c_2e^{r_2t} \tag{7.8}$$

donde las c son constantes que se determinan a partir de las condiciones iniciales. Este caso se llama *sobreamortiguado.*

Si el discriminante es cero, resulta una sola raíz real y la solución general se escribe como

$$y = (c_1 + c_2t)e^{\lambda t} \tag{7.9}$$

Este caso se llama *con amortiguamiento crítico.*

Si el discriminante es negativo, las raíces son números complejos conjugados

$$\begin{matrix} r_1 \\ r_2 \end{matrix} = \lambda \pm \mu i$$

y la solución general se formula como

$$y = c_1e^{(\lambda+\mu i)t} + c_2e^{(\lambda-\mu i)t}$$

El comportamiento fisíco de esta solución se aclara mediante la fórmula de Euler

$$e^{\mu i t} = \cos \mu t + i \text{ sen } \mu t$$

para reformular la solución general como (véase Boyce y DiPrima, 1992, para detalles de la demostración)

$$y = c_1 e^{\lambda t} \cos \mu t + c_2 e^{\lambda t} \text{ sen } \mu t \tag{7.10}$$

Este caso se llama *subamortiguado*.

Las ecuaciones (7.8), (7.9) y (7.10) expresan las maneras posibles en que los sistemas lineales responden dinámicamente. El término exponencial indica que la solución del sistema es capaz de decaer (parte real negativa) o crecer (parte real positiva) exponencialmente con el tiempo (figura 7.1*a*). El término senosoidal (parte imaginaria) significa que la solución puede oscilar (figura 7.1*b*). Si el eigenvalor tiene tanto parte real como imaginaria, se combinan la forma exponencial y senosoidal (figura 7.1*c*). Debido a que este conocimiento es el elemento clave para entender, diseñar y controlar el comportamiento de sistemas físicos, los polinomios característicos son muy importantes en ingeniería y en muchas ramas de la ciencia. Se analizará la dinámica de varios sistemas de ingeniería en las aplicaciones que se estudian en el capítulo 8.

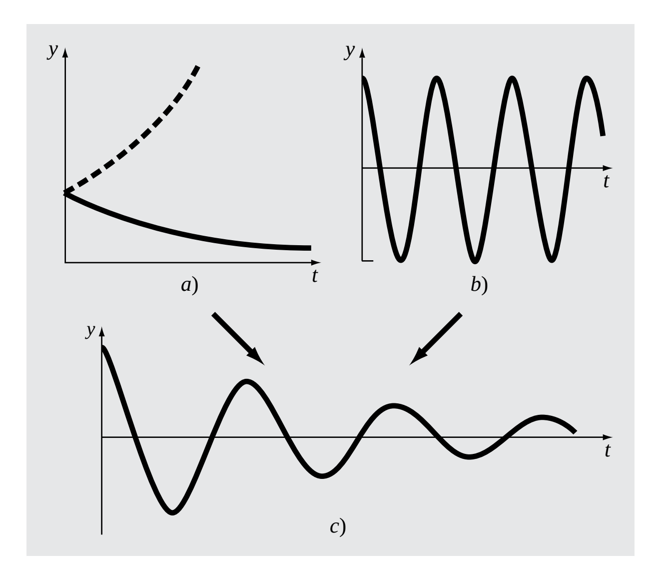


Figura 7.1
La solución general de las EDO lineales puede estar determinada por componentes *a*) exponenciales y *b*) senosoidales. La combinación de las dos formas es una senosoidal amortiguada como se muestra en *c*).

7.2 CÁLCULOS CON POLINOMIOS

Antes de describir los métodos para localizar raíces, se examinarán algunas operaciones fundamentales con polinomios. Dichas operaciones tendrán utilidad en sí mismas, además de proporcionar apoyo para localizar las raíces.

7.2.1 Evaluación y derivación de polinomios

Aunque la forma de la ecuación (7.1) es la más común, no resulta la mejor para determinar el valor de un polinomio para un valor específico de x. Por ejemplo, evaluar el polinomio de tercer grado como

$$f_3(x) = a_3x^3 + a_2x^2 + a_1x + a_0 \tag{7.11}$$

implica seis multiplicaciones y tres sumas. En general, para un polinomio de n-ésimo grado, se requieren $n(n + 1)/2$ multiplicaciones y n sumas.

La forma anidada, en cambio

$$f_3(x) = ((a_3x + a_2)x + a_1)x + a_0 \tag{7.12}$$

implica tres multiplicaciones y tres sumas. Para un polinomio de n-ésimo grado, esta forma requiere n multiplicaciones y n sumas. Ya que la forma anidada minimiza el número de operaciones, también tiende a minimizar los errores de redondeo. Observe que, según sea la preferencia, el orden de anidamiento puede invertirse:

$$f_3(x) = a_0 + x(a_1 + x(a_2 + xa_3)) \tag{7.13}$$

Un pseudocódigo adecuado para implementar la forma anidada se escribe simplemente como

```
DOFOR j = n, 0, -1
   p = p * x + a(j)
END DO
```

donde p tiene el valor del polinomio (definido por los coeficientes de las a) evaluado en x.

Existen casos (como el método de Newton-Raphson) donde se requiere evaluar tanto la función como su derivada. Esta evaluación se puede también incluir al agregar una línea en el pseudocódigo anterior,

```
DOFOR j = n, 0, -1
  df = df * x + p
  p = p * x + a(j)
END DO
```

donde *df* es la primera derivada del polinomio.

7.2.2 Deflación polinomial

Suponga que se determina una raíz de un polinomio de n-ésimo grado. Si se repite el procedimiento para localizar una raíz, puede encontrarse la misma raíz. Por lo tanto, sería adecuado eliminar la raíz encontrada antes de continuar. A este proceso de eliminar la raíz se le llama *deflación polinomial*.

Antes de mostrar cómo se hace esto, veamos algunos antecedentes útiles. Los polinomios son típicamente representados en la forma de la ecuación (7.1). Por ejemplo, un polinomio de quinto grado puede escribirse como

$$f_5(x) = -120 - 46x + 79x^2 - 3x^3 - 7x^4 + x^5 \tag{7.14}$$

Aunque ésta es la forma más común, no necesariamente es la mejor expresión para entender el comportamiento matemático de los polinomios. Por ejemplo, este polinomio de quinto grado se expresa de manera alternativa como

$$f_5(x) = (x + 1)(x - 4)(x - 5)(x + 3)(x - 2) \tag{7.15}$$

Ésta se conoce como la forma *factorizada* de un polinomio. Si se efectúa la multiplicación y se agrupan los términos semejantes, se obtendrá la ecuación (7.14). Sin embargo, la forma de la ecuación (7.15) tiene la ventaja de que indica claramente las raíces de la función. Así, resulta claro que $x = -1$, 4, 5, –3 y 2 son todas las raíces, porque cada una hace que uno de los términos de la ecuación (7.15) sea igual a cero.

Ahora, suponga que se divide este polinomio de quinto grado entre cualquiera de sus factores; por ejemplo, $x + 3$. En este caso, el resultado será un polinomio de cuarto grado

$$F_4(x) = (x + 1)(x - 4)(x - 5)(x - 2) = -40 - 2x + 27x^2 - 10x^3 + x^4 \tag{7.16}$$

con un residuo igual a cero.

En el pasado, quizás usted aprendió que los polinomios se dividen usando un procedimiento llamado *división sintética*. Varios algoritmos de computadora (basados tanto en la división sintética como en otros métodos) están disponibles para realizar la operación. Un esquema simple se proporciona en el siguiente pseudocódigo, el cual divide un polinomio de n-ésimo grado entre un factor monomial $x - t$.

```
r = a(n)
a(n) = 0
DOFOR i = n-1, 0, -1
  s = a(i)
  a(i) = r
  r = s + r * t
END DO
```

Si el monomio es una raíz del polinomio, el residuo r será cero, y los coeficientes del cociente se guardarán en a, al final del ciclo.

EJEMPLO 7.1 Deflación polinomial

Planteamiento del problema Divida el polinomio de segundo grado

$$f(x) = (x - 4)(x + 6) = x^2 + 2x - 24$$

entre el factor $x - 4$.

Solución Usando el método propuesto en el pseudocódigo anterior, los parámetros son $n = 2$, $a_0 = -24$, $a_1 = 2$, $a_2 = 1$ y $t = 4$. Estos valores se usan para calcular

$$r = a_2 = 1$$
$$a_2 = 0$$

El *loop* o ciclo se itera después desde $i = 2 - 1 = 1$ hasta 0. Para $i = 1$,

$$s = a_1 = 2$$
$$a_1 = r = 1$$
$$r = s + rt = 2 + 1(4) = 6$$

Para $i = 0$,

$$s = a_0 = -24$$
$$a_0 = r = 6$$
$$r = -24 + 6(4) = 0$$

Así, el resultado, como se esperaba, es el cociente $a_0 + a_1x = 6 + x$, con un residuo de cero.

También es posible dividir entre polinomios de grado superior. Como se verá más adelante en este capítulo, la tarea más común es dividir entre un polinomio de segundo grado o parábola. La subrutina de la figura 7.2 resuelve el problema más general de dividir un polinomio a de grado n entre un polinomio d de grado m. El resultado es un polinomio q de grado $(n - m)$, con un polinomio de grado $(m - 1)$ como el residuo.

```
SUB poldiv(a, n, d, m, q, r)
DOFOR j = 0, n
  r(j) = a(j)
  q(j) = 0
END DO
DOFOR k = n-m, 0, -1
  q(k+1) = r(m+k) / d(m)
  DOFOR j = m+k-1, k, -1
   r(j) = r(j)-q(k+1) * b(j-k)
  END DO
END DO
DOFOR j = m, n
  r(j) = 0
END DO
n = n-m
DOFOR i = 0, n
  a(i) = q(i+1)
END DO
END SUB
```

Figura 7.2
Algoritmo que divide un polinomio (definido por sus coeficientes a) entre un polinomio de grado menor d.

Ya que cada raíz calculada se conoce únicamente en forma aproximada, se observa que la deflación es sensible al error de redondeo. En algunos casos puede crecer a tal punto que los resultados lleguen a no tener sentido.

Algunas estrategias generales pueden aplicarse para minimizar el problema. Por ejemplo, el error de redondeo está afectado por el orden en que se evalúan los términos. La *deflación hacia adelante* se refiere al caso donde los coeficientes del nuevo polinomio están en grado de potencias descendentes de x (es decir, del término de mayor grado al de grado cero). En tal caso, es preferible dividir primero entre las raíces con el valor absoluto más pequeño. En forma inversa, en la *deflación hacia atrás* (esto es, del término de grado cero al de mayor grado) es preferible dividir primero entre las raíces con mayor valor absoluto.

Otra manera de reducir los errores de redondeo es considerar que cada raíz sucesiva estimada, obtenida durante la deflación es un buen primer valor inicial. Al utilizarse como un valor inicial, y determinar las raíces otra vez con el polinomio original sin deflación, se obtienen raíces que se conocen como *raíces pulidas*.

Por último, se presenta un problema cuando dos raíces deflacionadas son suficientemente inexactas, de tal manera que ambas convergen a la misma raíz no deflacionada. En tal caso, se podría creer en forma errónea que un polinomio tiene una raíz múltiple (recuerde la sección 6.5). Una forma para detectar este problema consiste en com-

parar cada raíz pulida con las que se han calculado anteriormente. Press *et al*. (2007) analizan el problema con mayor detalle.

7.3 MÉTODOS CONVENCIONALES

Ahora que se ha visto algún material de apoyo sobre polinomios, empezaremos a describir los métodos para localizar sus raíces. Es obvio que el primer paso sería investigar la posibilidad de usar los métodos cerrados y abiertos, descritos en los capítulos 5 y 6.

La eficacia de dichos métodos depende de que el problema a resolver tenga raíces complejas. Si sólo existen raíces reales, cualquiera de los métodos descritos anteriormente puede utilizarse. Sin embargo, el problema de encontrar un buen valor inicial complica tanto los métodos cerrados como los abiertos; además que los métodos abiertos podrían ser susceptibles a problemas de divergencia.

Cuando existen raíces complejas, los métodos cerrados obviamente no se pueden usar, ya que el criterio para definir el intervalo (que es el cambio de signo) no puede trasladarse a valores complejos.

De los métodos abiertos, el método convencional de Newton-Raphson llega a ofrecer una aproximación viable. En particular, es posible desarrollar un código conciso que comprenda deflación. Si se usa un lenguaje que permite manipular variables complejas (como Fortran), entonces el algoritmo localizará tanto raíces reales como complejas. Sin embargo, como es de esperarse, podría ser susceptible a tener problemas de convergencia. Por tal razón, se han desarrollado métodos especiales para encontrar raíces reales y complejas de polinomios. Se describen dos de estos métodos, el método de Müller y el de Bairstow, en las siguientes secciones. Como se verá, ambos están relacionados con los métodos abiertos convencionales descritos en el capítulo 6.

7.4 MÉTODO DE MÜLLER

Recuerde que el método de la secante obtiene una aproximación de la raíz dirigiendo una línea recta hasta el eje x con dos valores de la función (figura 7.3*a*). El método de Müller es similar; pero se construye una parábola con tres puntos (figura 7.3*b*).

El método consiste en obtener los coeficientes de la parábola que pasa por los tres puntos. Dichos coeficientes se sustituyen en la fórmula cuadrática para obtener el valor donde la parábola interseca al eje x; es decir, la raíz estimada. La aproximación se facilita al escribir la ecuación de la parábola en una forma conveniente,

$$f_2(x) = a(x - x_2)^2 + b(x - x_2) + c \tag{7.17}$$

Queremos que esta parábola pase por tres puntos $[x_0, f(x_0)]$, $[x_1, f(x_1)]$ y $[x_2, f(x_2)]$. Los coeficientes de la ecuación (7.17) se evalúan sustituyendo cada uno de esos tres puntos para dar

$$f(x_0) = a(x_0 - x_2)^2 + b(x_0 - x_2) + c \tag{7.18}$$

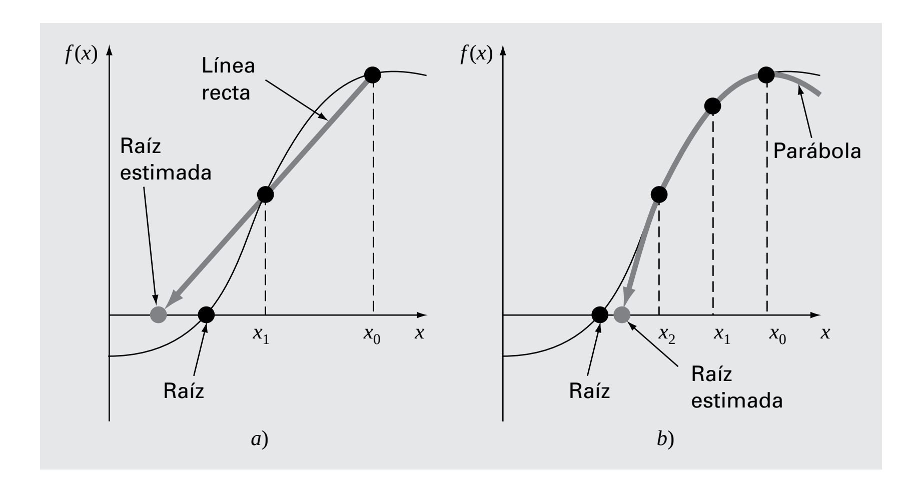


Figura 7.3
Una comparación de dos métodos relacionados para encontrar raíces: *a*) el método de la secante y *b*) el método de Müller.

$$f(x_1) = a(x_1 - x_2)^2 + b(x_1 - x_2) + c \tag{7.19}$$

$$f(x_2) = a(x_2 - x_2)^2 + b(x_2 - x_2) + c \tag{7.20}$$

Observe que se ha eliminado el subíndice "2" de la función por brevedad. Debido a que se tienen tres ecuaciones, es posible encontrar los tres coeficientes desconocidos a, b y c. Debido a que dos términos de la ecuación (7.20) son cero, se encuentra inmediatamente que $c = f(x_2)$. Así, el coeficiente c es igual al valor de la función evaluada en el tercer valor inicial, x_2. Este resultado se sustituye en las ecuaciones (7.18) y (7.19) para tener dos ecuaciones con dos incógnitas:

$$f(x_0) - f(x_2) = a(x_0 - x_2)^2 + b(x_0 - x_2) \tag{7.21}$$

$$f(x_1) - f(x_2) = a(x_1 - x_2)^2 + b(x_1 - x_2) \tag{7.22}$$

Una manipulación algebraica permite encontrar los coeficientes restantes a y b. La manera de hacer esto consiste en definir las diferencias:

$$h_0 = x_1 - x_0 \qquad h_1 = x_2 - x_1$$

$$\delta_0 = \frac{f(x_1) - f(x_0)}{x_1 - x_0} \qquad \delta_1 = \frac{f(x_2) - f(x_1)}{x_2 - x_1} \tag{7.23}$$

Éstas se sustituyen en las ecuaciones (7.21) y (7.22) para dar

$$(h_0 + h_1)b - (h_0 + h_1)^2 a = h_0\delta_0 + h_1\delta_1$$

$$h_1 \quad b - \quad h_1^2 \quad a = \quad h_1\delta_1$$

de donde se despejan a y b. El resultado se resume como

$$a = \frac{\delta_1 - \delta_0}{h_1 - h_0} \tag{7.24}$$

$$b = ah_1 + \delta_1 \tag{7.25}$$

$$c = f(x_2) \tag{7.26}$$

Para encontrar la raíz se aplica la fórmula cuadrática a la ecuación (7.17). Sin embargo, debido al error de redondeo potencial, en lugar de usar la forma convencional, se usará la fórmula alternativa [ecuación (3.13)], es decir,

$$x_3 - x_2 = \frac{-2c}{b \pm \sqrt{b^2 - 4ac}} \tag{7.27a}$$

o despejando la incógnita x_3 del lado izquierdo del signo igual,

$$x_3 = x_2 + \frac{-2c}{b \pm \sqrt{b^2 - 4ac}} \tag{7.27b}$$

Observe que al usar la fórmula cuadrática, es posible localizar tanto las raíces reales como las complejas. Ésta es la mayor ventaja del método.

Además, la ecuación (7.27*a*) proporciona una forma directa para determinar el error de aproximación. Debido a que el lado izquierdo representa la diferencia entre la raíz estimada actual (x_3) y la raíz estimada anterior (x_2), el error se calcula como

$$\varepsilon_a = \left| \frac{x_3 - x_2}{x_3} \right| 100\%$$

Ahora, un problema de la ecuación (7.27*a*) es que produce dos raíces, correspondientes a los términos ± del denominador. En el método de Müller, se escoge el signo que coincida con el signo de *b*. Esta elección proporciona como resultado el denominador más grande y, por lo tanto, dará la raíz estimada más cercana a x_2.

Una vez que se determinó x_3, el proceso se repite. Esto trae el problema de que un valor es descartado. En general, dos estrategias son comúnmente usadas:

1. Si sólo se localizan raíces reales, elegimos los dos valores originales más cercanos a la nueva raíz estimada, x_3.
2. Si se localizan raíces reales y complejas, se emplea un método secuencial. Es decir, como en el método de la secante, x_1, x_2 y x_3 toman el lugar de x_0, x_1 y x_2.

EJEMPLO 7.2 Método de Müller

Planteamiento del problema Utilice el método de Müller con valores iniciales x_0, x_1 y x_2 = 4.5, 5.5 y 5, respectivamente, para determinar la raíz de la ecuación

$$f(x) = x^3 - 13x - 12$$

Observe que las raíces de la ecuación son –3, –1 y 4.

Solución Primero se evaluará la función con los valores iniciales

$$f(4.5) = 20.625 \qquad f(5.5) = 82.875 \qquad f(5) = 48$$

que se emplean para calcular

$$h_0 = 5.5 - 4.5 = 1 \qquad h_1 = 5 - 5.5 = -0.5$$

$$\delta_0 = \frac{82.875 - 20.625}{5.5 - 4.5} = 62.25 \qquad \delta_1 = \frac{48 - 82.875}{5 - 5.5} = 69.75$$

Estos valores, a su vez, se sustituyen con las ecuaciones (7.24) a (7.26) para calcular

$$a = \frac{69.75 - 62.25}{-0.5 + 1} = 15 \qquad b = 15(-0.5) + 69.75 = 62.25 \qquad c = 48$$

La raíz cuadrada del discriminante se evalúa como

$$\sqrt{62.25^2 - 4(15)48} = 31.54461$$

Luego, como |62.25 + 31.54451| > |62.25 – 31.54451|, se emplea un signo positivo en el denominador de la ecuación (7.27*b*), y la nueva raíz estimada se determina como

$$x_3 = 5 + \frac{-2(48)}{62.25 + 31.54451} = 3.976487$$

y desarrollando el error estimado

$$\varepsilon_a = \left|\frac{-1.023513}{3.976487}\right| 100\% = 25.74\%$$

Debido a que el error es grande, se asignan nuevos valores: x_0 se reemplaza por x_1, x_1 se reemplaza por x_2 y x_2 se reemplaza por x_3. Por lo tanto, para la nueva iteración,

$$x_0 = 5.5 \qquad x_1 = 5 \qquad x_2 = 3.976487$$

y se repite el cálculo. Los resultados, tabulados a continuación, muestran que el método converge rápidamente a la raíz $x_r = 4$:

i	x_r	ε_a (%)
0	5	
1	3.976487	25.74
2	4.00105	0.6139
3	4	0.0262
4	4	0.0000119

El pseudocódigo del método de Müller para raíces reales se presenta en la figura 7.4. Observe que esta rutina toma un valor inicial único diferente de cero, que después se altera por el factor h para generar los otros dos valores iniciales. Por supuesto, el algoritmo puede programarse para considerarse tres valores iniciales. Con lenguajes parecidos a Fortran, el programa encontrará raíces complejas si las variables adecuadas se declaran como complejas.

```
SUB Müller(xr, h, eps, maxit)
x2 = xr
x1 = xr + h*xr
x0 = xr - h*xr
DO
  iter = iter + 1
  h0 = x1 - x0
  h1 = x2 - x1
  d0 = (f(x1) - f(x0)) / h0
  d1 = (f(x2) - f(x1)) / h1
  a = (d1 - d0) /(h1 + h0)
  b = a*h1 + d1
  c = f(x2)
  rad = SQRT(b*b - 4*a*c)
  If |b+rad| > |b-rad| THEN
    den = b + rad
  ELSE
    den = b - rad
  END IF
  dxr = -2*c /den
  xr = x2 + dxr
  PRINT iter, xr
  IF (|dxr| < eps*xr OR iter >= maxit) EXIT
  x0 = x1
  x1 = x2
  x2 = xr
END DO
END Müller
```

Figura 7.4
Pseudocódigo para el método de Müller.

7.5 MÉTODO DE BAIRSTOW

El método de Bairstow es un método iterativo relacionado de alguna manera con los métodos de Müller y de Newton-Raphson. Antes de hacer la descripción matemática de éste, recuerde la forma factorizada de un polinomio, por ejemplo

$$f_5(x) = (x + 1)(x - 4)(x - 5)(x + 3)(x - 2) \tag{7.28}$$

Si se divide entre un factor que no es una raíz (por ejemplo, $x + 6$), el cociente es un polinomio de cuarto grado. Aunque, en este caso, habrá un residuo.

Con estas consideraciones se puede elaborar un algoritmo para determinar la raíz de un polinomio: 1) dé un valor inicial para la raíz $x = t$; 2) divida el polinomio entre el factor $x - t$, y 3) deter-

mine si hay un residuo diferente de cero. Si no, el valor inicial es perfecto y la raíz es igual a t. Si existe un residuo, se ajusta el valor inicial en forma sistemática y se repite el procedimiento hasta que el residuo desaparezca y se localice la raíz. Una vez hecho esto, se repite el procedimiento totalmente, ahora con el cociente para localizar otra raíz.

Por lo general, el método de Bairstow se basa en esta manera de proceder. Por consiguiente, depende del proceso matemático de dividir un polinomio entre un factor. Recuerde (sección 7.2.2) de nuestro estudio de la deflación de polinomios que la división sintética implica la división del polinomio entre un factor $x - t$. Por ejemplo, el polinomio general [ecuación (7.1)]

$$f_n(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \tag{7.29}$$

se divide entre el factor $x - t$ para dar un segundo polinomio que es de un grado menor:

$$f_{n-1}(x) = b_1 + b_2x + b_3x^2 + \cdots + b_nx^{n-1} \tag{7.30}$$

con un residuo $R = b_0$, donde los coeficientes se calculan por la relación de recurrencia

$$b_n = a_n$$

$$b_i = a_i + b_{i+1}t \qquad \text{para } i = n - 1 \text{ a } 0$$

Observe que si t es una raíz del polinomio original, el residuo b_0 sería igual a cero.

Para permitir la evaluación de raíces complejas, el método de Bairstow divide el polinomio entre un factor cuadrático $x^2 - rx - s$. Si esto se hace con la ecuación (7.29), el resultado es un nuevo polinomio

$$f_{n-2}(x) = b_2 + b_3x + \cdots + b_{n-1}x^{n-3} + b_nx^{n-2}$$

con un residuo

$$R = b_1(x - r) + b_0 \tag{7.31}$$

Como con la división sintética normal, se utiliza una relación de recurrencia simple para realizar la división entre el factor cuadrático:

$$b_n = a_n \tag{7.32a}$$

$$b_{n-1} = a_{n-1} + rb_n \tag{7.32b}$$

$$b_i = a_i + rb_{i+1} + sb_{i+2} \quad \text{para } i = n - 2 \text{ a } 0 \tag{7.32c}$$

El factor cuadrático se introduce para permitir la determinación de las raíces complejas. Esto se relaciona con el hecho de que, si los coeficientes del polinomio original son reales, las raíces complejas se presentan en pares conjugados. Si $x^2 - rx - s$ es un divisor exacto del polinomio, las raíces complejas pueden determinarse con la fórmula cuadrática. Así, el método se reduce a determinar los valores de r y s que hacen que el factor cuadrático sea un divisor exacto. En otras palabras, se buscan los valores que hacen que el residuo sea igual a cero.

La inspección de la ecuación (7.31) nos lleva a concluir que para que el residuo sea cero, b_0 y b_1 deben ser cero. Como es improbable que los valores iniciales para evaluar r y s conduzcan a este resultado, debemos determinar una forma sistemática para modificar los valores iniciales, de tal forma que b_0 y b_1 tiendan a cero. Para lograrlo, el método de Bairstow usa una estrategia similar a la del método de Newton-Raphson. Como tanto b_0 como b_1 son funciones de r y s, se pueden expandir usando una serie de Taylor, así [recuerde la ecuación (4.26)]:

$$b_1(r + \Delta r, s + \Delta s) = b_1 + \frac{\partial b_1}{\partial r}\Delta r + \frac{\partial b_1}{\partial s}\Delta s$$

$$b_0(r + \Delta r, s + \Delta s) = b_0 + \frac{\partial b_0}{\partial r}\Delta r + \frac{\partial b_0}{\partial s}\Delta s \tag{7.33}$$

donde los valores del lado derecho se evalúan en r y s. Observe que se han despreciado los términos de segundo grado y de grado superior. Esto representa una suposición implícita de que $-r$ y $-s$ son

suficientemente pequeños para que los términos de orden superior puedan despreciarse. Otra manera de expresar esta suposición es que los valores iniciales son adecuadamente cercanos a los valores de r y s en las raíces.

Los incrementos, Δr y Δs, necesarios para mejorar nuestros valores iniciales, se estiman igualando a cero la ecuación (7.33) para dar

$$\frac{\partial b_1}{\partial r}\Delta r + \frac{\partial b_1}{\partial s}\Delta s = -b_1 \tag{7.34}$$

$$\frac{\partial b_0}{\partial r}\Delta r + \frac{\partial b_0}{\partial s}\Delta s = -b_0 \tag{7.35}$$

Si las derivadas parciales de las b pueden determinarse, hay un sistema de dos ecuaciones que se resuelve simultáneamente para las dos incógnitas, Δr y Δs. Bairstow demostró que las derivadas parciales se obtienen por división sintética de las b en forma similar a como las b mismas fueron obtenidas:

$$c_n = b_n \tag{7.36a}$$

$$c_{n-1} = b_{n-1} + rc_n \tag{7.36b}$$

$$c_i = b_i + rc_{i+1} + sc_{i+2} \qquad \text{para } i = n-2 \text{ a } 1 \tag{7.36c}$$

donde $\partial b_0/\partial r = c_1$, $\partial b_0/\partial s = \partial b_1/\partial r = c_2$ y $\partial b_1/\partial s = c_3$. Así, las derivadas parciales se obtienen por la división sintética de las b. Entonces, las derivadas parciales se sustituyen en las ecuaciones (7.34) y (7.35) junto con las b para dar

$$c_2\Delta r + c_3\Delta s = -b_1$$

$$c_1\Delta r + c_2\Delta s = -b_0$$

Estas ecuaciones se resuelven para Δr y Δs, las cuales, a su vez, se emplean para mejorar los valores iniciales de r y s. En cada paso, se estima un error aproximado en r y s:

$$|\varepsilon_{a,r}| = \left|\frac{\Delta r}{r}\right| 100\% \tag{7.37}$$

y

$$|\varepsilon_{a,s}| = \left|\frac{\Delta s}{s}\right| 100\% \tag{7.38}$$

Cuando ambos errores estimados caen por debajo de un criterio especificado de terminación ε_s, los valores de las raíces se determinan mediante

$$x = \frac{r \pm \sqrt{r^2 + 4s}}{2} \tag{7.39}$$

En este punto, existen tres posibilidades:

1. *El cociente es un polinomio de tercer grado o mayor*. En tal caso, el método de Bairstow se aplica al cociente para evaluar un nuevo valor de r y s. Los valores anteriores de r y s pueden servir como valores iniciales en esta aplicación.
2. *El cociente es cuadrático.* Aquí es posible evaluar directamente las dos raíces restantes con la ecuación (7.39).
3. *El cociente es un polinomio de primer grado.* En este caso, la raíz restante se evalúa simplemente como

$$x = -\frac{s}{r} \tag{7.40}$$

EJEMPLO 7.3 Método de Bairstow

Planteamiento del problema Emplee el método de Bairstow para determinar las raíces del polinomio

$$f_5(x) = x^5 - 3.5x^4 + 2.75x^3 + 2.125x^2 - 3.875x + 1.25$$

Utilice como valores iniciales $r = s = -1$ e itere hasta un nivel de $\varepsilon_s = 1\%$.

Solución Se aplican las ecuaciones (7.32) y (7.36) para calcular

$b_5 = 1$ $b_4 = -4.5$ $b_3 = 6.25$ $b_2 = 0.375$ $b_1 = -10.5$

$b_0 = 11.375$

$c_5 = 1$ $c_4 = -5.5$ $c_3 = 10.75$ $c_2 = -4.875$ $c_1 = -16.375$

Así, las ecuaciones simultáneas para encontrar Δr y Δs son

$$-4.875\Delta r + 10.75\Delta s = 10.5$$
$$-16.375\Delta r - 4.875\Delta s = -11.375$$

al ser resueltas se encuentra que $\Delta r = 0.3558$ y $\Delta s = 1.1381$. Por lo tanto, nuestros valores iniciales se corrigen a

$$r = -1 + 0.3558 = -0.6442$$
$$s = -1 + 1.1381 = 0.1381$$

y se evalúa el error aproximado con las ecuaciones (7.37) y (7.38),

$$|\varepsilon_{a,r}| = \left|\frac{0.3558}{-0.6442}\right| 100\% = 55.23\% \qquad |\varepsilon_{a,s}| = \left|\frac{1.1381}{0.1381}\right| 100\% = 824.1\%$$

A continuación, se repiten los cálculos usando los valores revisados para r y s. Aplicando las ecuaciones (7.32) y (7.36) se obtiene

$b_5 = 1$ $b_4 = -4.1442$ $b_3 = 5.5578$ $b_2 = -2.0276$ $b_1 = -1.8013$
$b_0 = 2.1304$
$c_5 = 1$ $c_4 = -4.7884$ $c_3 = 8.7806$ $c_2 = -8.3454$ $c_1 = 4.7874$

Por lo tanto, se debe resolver el sistema de ecuación

$$-8.3454\Delta r + 8.7806\Delta s = 1.8013$$
$$4.7874\Delta r - 8.3454\Delta s = -2.1304$$

al tener la solución $\Delta r = 0.1331$ y $\Delta s = 0.3316$, ésta se utiliza para corregir la raíz estimada:

$r = -0.6442 + 0.1331 = -0.5111$ $|\varepsilon_{a,r}| = 26.0\%$

$s = 0.1381 + 0.3316 = 0.4697$ $|\varepsilon_{a,s}| = 70.6\%$

El cálculo continúa, resultando que después de cuatro iteraciones el método converge a los valores $r = -0.5$ ($|\varepsilon_{a,r}| = 0.063\%$) y $s = 0.5$ ($|\varepsilon_{a,s}| = 0.040\%$). La ecuación (7.39) puede emplearse para evaluar las raíces:

$$x = \frac{-0.5 \pm \sqrt{(-0.5)^2 + 4(0.5)}}{2} = 0.5, -1.0$$

Entonces, se tiene que el cociente es la ecuación cúbica

$$f(x) = x^3 - 4x^2 + 5.25x - 2.5$$

El método de Bairstow puede aplicarse a este polinomio usando los resultados del paso anterior, $r = -0.5$ y $s = 0.5$, como valores iniciales. Cinco iteraciones dan las aproximaciones $r = 2$ y $s = -1.249$, las cuales se usan para calcular

$$x = \frac{2 \pm \sqrt{2^2 + 4(-1.249)}}{2} = 1 \pm 0.499i$$

Ahora, el cociente es un polinomio de primer grado que puede ser directamente evaluado mediante la ecuación (7.40) para determinar la quinta raíz: 2.

Observe que la esencia del método de Bairstow es la evaluación de las b y de las c por medio de las ecuaciones (7.32) y (7.36). Una de las ventajas principales de este método radica en la forma concisa en la cual tales fórmulas de recurrencia pueden programarse.

***a*) Algoritmo de Bairstow**

```
SUB Bairstow (a,nn,es,rr,ss,maxit,re,im,ier)
DIMENSION b(nn), c(nn)
r = rr; s = ss; n = nn
ier = 0; ea1 = 1; ea2 = 1
DO
  IF n < 3 OR iter ≥ maxit EXIT
  iter = 0
  DO
    iter = iter + 1
    b(n) = a(n)
    b(n - 1) = a(n - 1) + r * b(n)
    c(n) = b(n)
    c(n - 1) = b(n - 1) + r * c(n)
    DO i = n - 2, 0, -1
      b(i) = a(i) + r * b(i + 1) + s * b(i + 2)
      c(i) = b(i) + r * c(i + 1) + s * c(i + 2)
    END DO
    det = c(2) * c(2)-c(3) *c(1)
    IF det ≠ 0 THEN
      dr = (-b(1) * c(2) + b(0) * c(3))/det
      ds = (-b(0) * c(2) + b(1) * c(1))/det
      r = r + dr
      s = s + ds
      IF r≠0 THEN ea1 = ABS(dr/r) * 100
      IF s≠0 THEN ea2 = ABS(ds/s) * 100
    ELSE
      r = r + 1
      s = s + 1
      iter = 0
    END IF
    IF ea1 ≤ es AND ea2 ≤ es OR iter ≥ maxit EXIT
  END DO
  CALL Quadroot(r,s,r1,i1,r2,i2)
  re(n) = r1
  im(n) = i1
  re(n - 1) = r2
  im(n - 1) = i2
  n = n-2
  DO i = 0, n
    a(i) = b(i + 2)
  END DO
END DO
IF iter < maxit THEN
  IF n = 2 THEN
    r = -a(1)/a(2)
    s = -a(0)/a(2)
    CALL Quadroot(r,s,r1,i1,r2,i2)
    re(n) = r1
    im(n) = i1
    re(n - 1) = r2
    im(n - 1) = i2
  ELSE
    re(n) = -a(0)/a(1)
    im(n) = 0
  END IF
ELSE
  ier = 1
END IF
END Bairstow
```

***b*) Algoritmo para raíces de una cuadrática**

```
SUB Quadroot(r,s,r1,i1,r2,i2)
disc = r ^ 2 + 4 * s
IF disc > 0 THEN
  r1 = (r + SQRT(disc))/2
  r2 = (r - SQRT(disc))/2
  i1 = 0
  i2 = 0
ELSE
  r1 = r/2
  r2 = r1
  i1 = SQRT(ABS(disc))/2
  i2 = -i1
END IF
END QuadRoot
```

Figura 7.5
a) Algoritmo para el método de Bairstow junto con *b*) un algoritmo para determinar las raíces de una ecuación cuadrática.

En la figura 7.5 se muestra el pseudocódigo que ejecuta el método de Bairstow. La parte principal de este algoritmo es el ciclo que evalúa las *b* y *c*. También observe que el pseudocódigo para resolver las ecuaciones simultáneas revisa para evitar la división entre cero. Si éste es el caso, los valores de *r* y *s* se alteran ligeramente y el procedimiento comienza de nuevo. Además, en el algoritmo hay un lugar donde el usuario puede definir el número máximo de iteraciones (MAXIT) y está diseñado para evitar una división entre cero cuando se calcula el error estimado. Finalmente, el algoritmo requiere valores iniciales para *r* y *s* (*rr* y *ss* en el código). Si no se tiene conocimiento *a priori* de que existan las raíces, se tendrá un conjunto de ceros al llamar el programa.

7.6 OTROS MÉTODOS

Otros métodos están disponibles para localizar las raíces de los polinomios. El *método de Jenkins-Traub* es comúnmente usado en bibliotecas de software. Es, en cierta medida, complicado y un punto de partida aceptable para entenderlo se encuentra en Ralston y Rabinowitz (1978).

El *método de Laguerre*, que aproxima las raíces reales y complejas, tiene una convergencia cúbica, se encuentra entre los mejores métodos. Un análisis completo se encuentra en Householder (1970). Además, Press *et al*. (1992) ofrecen un buen algoritmo para implementar este método.

7.7 LOCALIZACIÓN DE RAÍCES CON PAQUETES DE SOFTWARE

Los paquetes de software tienen gran capacidad para localizar raíces. En esta sección, se ofrece una muestra de los más útiles.

7.7.1 Excel

Una hoja de cálculo como Excel se utiliza para localizar la raíz mediante *prueba y error.* Por ejemplo, si se quiere encontrar una raíz de

$$f(x) = x - \cos x$$

primero se introduce un valor de x en una celda. Después se destina otra celda para $f(x)$ donde se obtendrá el valor de la función para la x de la primera celda. Se puede variar el valor de la celda en x hasta que la celda de $f(x)$ se aproxime a cero. Este proceso se mejora usando la capacidad de graficación de Excel para obtener un buen valor inicial (figura 7.6).

Aunque Excel facilita el método de prueba y error, también posee dos herramientas estándar que sirven para la localización de raíces: *Goal Seek* (buscar objetivo) y *Solver.* Ambas son útiles para ajustar sistemáticamente los valores iniciales. Goal Seek se utiliza expresamente para llevar la ecuación a un valor (en este caso, cero) mediante la variación de un solo parámetro.

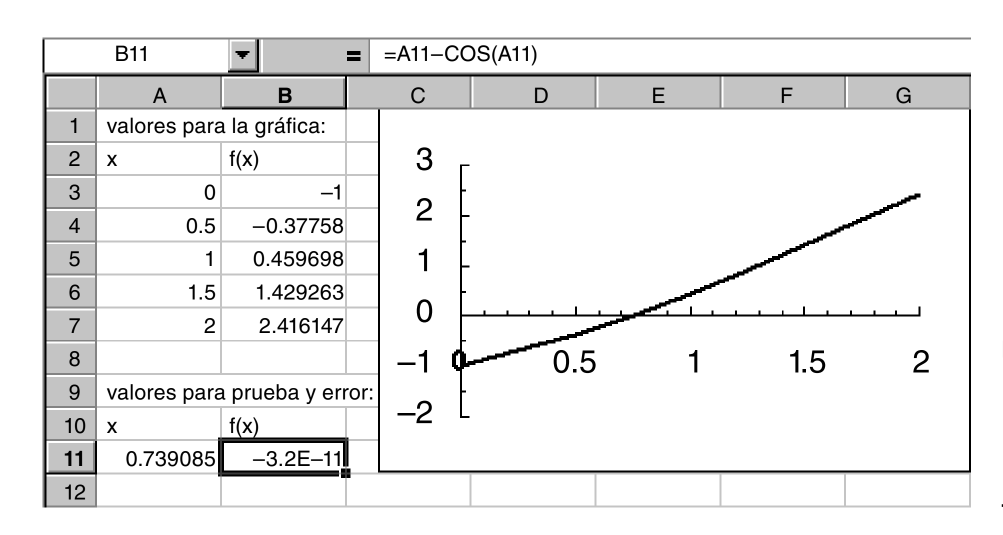


Figura 7.6
Una hoja de cálculo para determinar la raíz de $f(x) = x - \cos x$ por prueba y error. La gráfica se usa para obtener un buen valor inicial.

EJEMPLO 7.4 Use la herramienta Goal Seek (buscar objetivo) de Excel para localizar una raíz simple

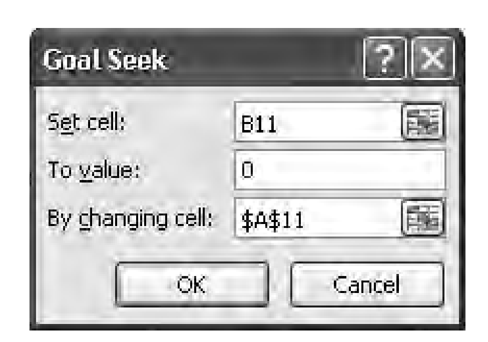
Planteamiento del problema Emplee "buscar objetivo" para determinar la raíz de la función trascendente

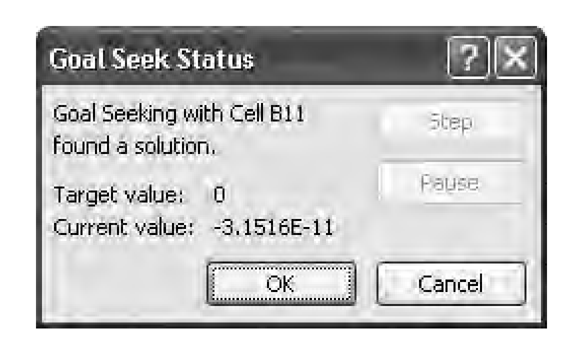
$$f(x) = x - \cos x$$

Solución Como en la figura 7.6, la clave para resolver una sola ecuación con Excel es crear una celda que tenga el valor de la función en cuestión y hacer, después, el valor dependiente de otra celda. Una vez hecho esto, de la barra de herramientas se selecciona el menú Data (datos), después, dentro de What-if-Analysis, se selecciona "Goal Seek". Ahora aparece una ventana de diálogo pidiendo se especifique una celda para un valor al modificar otra celda. Por ejemplo, suponga que, como en la figura 7.6, el valor propuesto se escribe en la celda A11 y la función resultante en la celda B11. La ventana de diálogo para Goal Seek será la de la derecha.

Cuando se selecciona el botón de OK (aceptar) una ventana de mensaje presenta los resultados de la ventana de abajo.

Las celdas de la hoja de cálculo se modificarán con los nuevos valores, como se muestra en la figura 7.6.





La herramienta *Solver* es más sofisticada que Goal Seek porque 1) puede variar simultáneamente varias celdas y 2) además de llevar la celda destino a un valor, éste puede minimizarse o maximizarse. En el siguiente ejemplo se ilustra cómo se utiliza para resolver un sistema de ecuaciones no lineales.

EJEMPLO 7.5 Uso de Solver de Excel para un sistema no lineal

Planteamiento del problema En la sección 6.6 obtuvimos la solución del siguiente sistema de ecuaciones simultáneas:

$$u(x, y) = x^2 + xy - 10 = 0$$

$$v(x, y) = y + 3xy^2 - 57 = 0$$

Observe que un par de raíces es $x = 2$ y $y = 3$. Utilice Solver para determinar las raíces usando como valores iniciales $x = 1$ y $y = 3.5$.

Solución Como se muestra en la siguiente tabla, dos celdas (B1 y B2) pueden crearse para los valores iniciales x y y. Los valores de la función, $u(x, y)$ y $v(x, y)$, pueden entrar en otras celdas (B3 y B4). Como se observa, los valores iniciales dan como resultado valores de la función que son lejanos a cero.

Después, se crea otra celda que contenga un valor que refleje qué tan cercanas de cero están ambas funciones. Una forma de

B6 | f_x =B3^2+B4^2

	A	B	C
1	x	1	
2	y	3.5	
3	u (x, y)	–5.5	
4	v(x, y)	–16.75	
5			
6	Suma de cuadrados	310.8125	
7			

hacerlo consiste en sumar los cuadrados de los valores de las funciones. Este resultado se introduce en la celda B6. Si ambas funciones son cero, esta función deberá también ser cero. Además, usando los cuadrados de las funciones se evita la posibilidad de que ambas funciones puedan tener el mismo valor diferente de cero, pero con signos contrarios. En tal caso, la celda de apoyo (B6) podría ser cero, aunque las raíces podrían ser incorrectas.

Una vez que la hoja de cálculo ha sido creada, se elige la opción **Solver** de la tira de datos.[1] Entonces, una ventana de diálogo se presentará en pantalla, pidiéndole la información pertinente. Las celdas solicitadas en la ventana de diálogo de Solver se llenarán como



	A	B	C	D
1	x	2.00003		
2	y	2.999984		
3	u(x, y)	0.000176		
4	v(x, y)	0.000202		
5				
6	Suma de cuadrados	7.19E-08		
7				

Cuando el botón de OK (aceptar) se selecciona, se abrirá una ventana de diálogo con un reporte de las operaciones efectuadas. En el presente caso, Solver obtiene la solución correcta:

Se debe observar que Solver puede fallar. Su éxito depende de 1) la condición del sistema de ecuaciones y/o 2) la calidad de los valores iniciales. El resultado satisfactorio del ejemplo anterior no está garantizado. A pesar de esto, se puede encontrar a Solver bastante útil para hacer de él una buena opción en la obtención rápida de raíces para un amplio rango de aplicaciones a la ingeniería.

7.7.2 MATLAB

Como se resume en la tabla 7.1, MATLAB es capaz de localizar raíces en ecuaciones algebraicas y trascendentes. Es excelente para manipular y localizar raíces en los polinomios.

La función `fzero` está diseñada para localizar la raíz de una función. Una representación simplificada de su sintaxis es

```
fzero(f,x0,opciones)
```

[1] Observe que usted puede necesitar instalar Solver, eligiendo entre las opciones de Excel, Office, Adiciones (Add-Ins), seleccionar Excel Add-Ins de la casilla de tomar y descargar (Drop Down box) en la parte inferior del Menú de Opciones Excel (Excel Options), y haga clic en Ir (Go). Luego ponga paloma en la casilla Solver. Así debe quedar instalado el Solver, y debe aparecer en su tira de Datos el respectivo botón de acceso.

donde `f` es la función que se va a analizar, x_0 es el valor inicial y *`opciones`* son los parámetros de optimización (éstos pueden cambiarse al usar la función `optimset`). Si no se anotan las opciones se emplean los valores por omisión. Observe que se pueden emplear uno o dos valores iniciales, asumiendo que la raíz está dentro del intervalo. El siguiente ejemplo ilustra cómo se usa la función `fzero`.

Tabla 7.1 Funciones comunes de MATLAB relacionadas con la manipulación de polinomios y la localización de raíces.

Función	Descripción
`fzero`	Raíz de una sola función
`roots`	Encuentra raíces de polinomios
`poly`	Construye polinomios con raíces específicas
`polival`	Evalúa un polinomio
`polivalm`	Evalúa un polinomio con argumento matricial
`residue`	Expansión de la fracción-parcial (residuos)
`polyder`	Diferenciación polinomial
`conv`	Multiplicación de polinomios
`deconv`	División de polinomios

EJEMPLO 7.6 Uso de MATLAB para localizar raíces

Planteamiento del problema Utilice la función `fzero` de MATLAB para encontrar las raíces de

$$f(x) = x^{10} - 1$$

dentro del intervalo $x_l = 0$ y $x_u = 4$, obviamente se tiene dos raíces –1 y 1. Recuerde que para determinar la raíz positiva en el ejemplo 5.6 se usó el método de la falsa posición con valores iniciales 0 y 1.3.

Solución Bajo las mismas condiciones iniciales del ejemplo 5.6, se usa MATLAB para determinar la raíz positiva

```
>> x0=[0 1.3];
>> x=fzero(@(x) x^10–1,x0)

x =
    1
```

De manera semejante, se emplean los valores iniciales –1.3 y 0 para determinar la raíz negativa

```
>> x0=[–1.3 0];
>> x=fzero(@(x) x^10–1,x0)

x =
    –1
```

Se puede usar un valor único; resulta un caso interesante cuando se usa el valor inicial 0

```
>> x0=0;
>> x=fzero(@(x) x^10–1,x0)

x =
    –1
```

Se tiene que para ese valor el algoritmo llevará a la raíz a su valor negativo.

El uso de `optimset` se ilustra al mostrar en pantalla la forma en que las iteraciones conducen a la solución

```
>> x0=0;
>> option=optimset('DISP','ITER');
>> x=fzero(@(x) x^10–1,x0,option)
```

SOFTWARE

```
Func–count    x            f(x)          Procedure

    1           0              –1      initial
    2   –0.0282843             –1      search
    3    0.0282843             –1      search
    4       –0.04              –1      search
    •
    •
    •
   21        0.64       –0.988471      search
   22   –0.905097       –0.631065      search
   23    0.905097       –0.631065      search
   24       –1.28         10.8059      search

Looking for a zero in the interval [–1.28], 0.9051]

   25    0.784528       –0.911674      interpolation
   26   –0.247736       –0.999999      bisection
   27   –0.763868       –0.932363      bisection
   28    –1.02193        0.242305      bisection
   29   –0.968701        –0.27239      interpolation
   30   –0.996873      –0.0308299      interpolation
   31   –0.999702     –0.00297526      interpolation
   32          –1     5.53132e-006     interpolation
   33          –1    –7.41965e-009     interpolation
   34          –1    –1.88738e-014     interpolation
   35          –1               0      interpolation
Zero found in the interval: [–1.28, 0.9051].

x =
    –1
```

Estos resultados ilustran la estrategia empleada por `fzero` cuando se tiene un valor único. Primero busca en la vecindad del valor inicial hasta detectar un cambio de signo. Después usa una combinación del método de bisección e interpolación para dirigirse a la raíz. La interpolación considera tanto el método de la secante como la interpolación cuadrática inversa (recuerde la sección 7.4). Deberá notar que el algoritmo de fzero puede implicar más cosas a partir de esta descripción básica. Puede consultar a Press *et al*. (2007) para mayores detalles.

EJEMPLO 7.7 Uso de MATLAB para manipular y determinar las raíces de polinomios

Planteamiento del problema Analicemos cómo se emplea MATLAB para manipular y determinar las raíces de polinomios. Use la siguiente ecuación del ejemplo 7.3,

$$f_5(x) = x^5 - 3.5x^4 + 2.75x^3 + 2.125x^2 - 3.875x + 1.25 \quad \text{(E7.7.1)}$$

que tiene tres raíces reales: 0.5, –1.0, 2 y un par de raíces complejas: $1 \pm 0.5i$.

Solución El polinomio se introduce en MATLAB almacenando los coeficientes como un vector. Por ejemplo, respondiendo a la instrucción en pantalla de MATLAB (>>), teclear e introducir la siguiente línea almacena los coeficientes en el vector `a`,

```
>> a=[1 –3.5 2.75 2.125 –3.875 1.25];
```

Después se procede a manipular el polinomio. Por ejemplo, podemos evaluarlo en $x = 1$, tecleando

```
>> polival(a,1)
```

que resultará $1(1)^5 - 3.5(1)^4 + 2.75(1)^3 + 2.125(1)^2 - 3.875(1) + 1.25 = -0.25$,

```
ans =
   -0.2500
```

Para evaluar la derivada $f'(x) = 5x^4 - 14x^3 + 8.25x^2 + 4.25x - 3.875$ con

```
>> polyder (a)
ans =
      5.0000     -14.0000     8.2500     4.2500     -3.8750
```

A continuación, se crea un polinomio cuadrático que tiene dos de las raíces originales de la ecuación (E7.7.1): 0.5 y –1. Esta cuadrática es $(x - 0.5)(x + 1) = x^2 + 0.5x - 0.5$ y se introduce en MATLAB como el vector b

```
>> b=[1 0.5 -0.5];
```

Se divide el polinomio original entre este polinomio con

```
>> [d, e]=deconv(a, b)
```

El resultado de la división es (un polinomio de tercer grado d) y un residuo (e)

```
d =
    1.0000   -4.0000   5.2500   -2.5000
e =
    0     0     0     0     0     0
```

Debido a que el polinomio es un divisor perfecto, el residuo polinominal tiene coeficientes iguales a cero. Ahora las raíces del cociente polinominal se determinan como

```
>> roots(d)
```

con el resultado esperado para las raíces faltantes del polinomio original (E7.7.1)

```
ans =
      2.0000
      1.0000 + 0.5000i
      1.0000 - 0.5000i
```

Ahora al multiplicar d por b se regresa al polinomio original

```
>> conv(d, b)
ans =
      1.0000   -3.5000   2.7500   2.1250   -3.8750   1.2500
```

Finalmente, podemos determinar todas las raíces del polinomio original con

```
>> r=roots(a)
r =
    -1.0000
     2.0000
     1.0000 + 0.5000i
     1.0000 - 0.5000i
     0.5000
```

7.7.3 Mathcad

Mathcad tiene una función de modo numérico llamada ***root*** que es posible usar para resolver una ecuación de una sola variable. El método exige que usted suministre una función $f(x)$, ya sea un valor inicial o un rango. Cuando se usa un solo valor inicial, **root** emplea los métodos *Secant* (o de la secante) y *Müller.* En caso de que se suministren dos valores iniciales entre los cuales está comprendida una raíz, utiliza una combinación del *método Ridder* (una variación del método de posición falsa) y del método de *Brent*. Itera hasta que la magnitud de $f(x)$ en la raíz propuesta es menor que el valor predefinido de **TOL**. La implementación de Mathcad tiene ventajas y desventajas similares que los métodos convencionales de determinación de raíces, como lo que concierne a la calidad de los valores iniciales y la velocidad de convergencia.

Mathcad puede encontrar todas las raíces reales o complejas de polinomios con ***polyroots***. Esta función de modo numérico o simbólico se basa en el *método de Laguerre*. Esta función no necesita valores iniciales, y todas las raíces las devuelve el sistema al mismo tiempo.

Mathcad contiene una función de modo numérico que se llama ***Find***, que se puede usar para resolver hasta 50 ecuaciones algebraicas simultáneas no lineales. La función **Find** elige un método adecuado de un grupo de métodos disponibles, dependiendo de si el problema es lineal o no lineal, y de otros atributos. Los valores aceptables para la solución pueden ser no restringidos o restringidos a caer dentro de límites especificados. Si **Find** no logra encontrar una solución que satisfaga las ecuaciones y las restricciones, produce un mensaje de error: "Did not find solution". Sin embargo, Mathcad también contiene una función similar que se llama **Minerr**. Esta función da resultados de solución que minimizan el error en las restricciones, incluso en el caso de no poder encontrar soluciones exactas. Así, el problema de despejar las raíces de ecuaciones no lineales está estrechamente relacionado tanto con la optimización como con los mínimos cuadrados no lineales. Estas áreas y el **Minerr** se analizan a detalle en las partes cuatro y cinco.

La figura 7.7 muestra una típica hoja de trabajo Mathcad. Los menús en la parte superior proporcionan rápido acceso a los operadores y a funciones aritméticas comunes y a las funciones, a varios tipos de gráficos de dos y tres dimensiones, y el ambiente para crear subprogramas. Se pueden colocar ecuaciones, texto, datos o gráficas en cualquier parte de la pantalla. Usted puede usar diversos tipos de impresión, colores y estilos para construir hojas de trabajo con casi cualquier diseño y formato que le agrade. Consulte el resumen del Manual del Usuario Mathcad (Mathcad User's Manual) en el apéndice C, o el manual completo disponible en MathSoft. Observe que en todos los ejemplos de Mathcad hemos tratado de acomodar la sesión completa de Mathcad en una sola pantalla. Usted debe percatarse de que la gráfica tendría que colocarse debajo de los comandos para que funcione correctamente.

Se comenzará con un ejemplo que despeja la raíz de $f(x) = x - \cos x$. El primer paso es introducir la función. Esto se hace tecleando f(x):, que Mathcad convierte automáticamente en f(x): =. La expresión := se llama símbolo de definición. En seguida se introduce un valor inicial de manera similar, usando el símbolo de definición. Ahora, la solución se define como **root**(f(x), x), que abre el método de secante con un valor inicial de 1.0. Se continúa la iteración hasta que $f(x)$ evaluada en la raíz propuesta sea menor que **TOL**. El valor de **TOL** se ajusta a partir del menú desplegable de opciones matemáticas (Math/Options). Finalmente se muestra en pantalla el valor de la solución (soln) usando

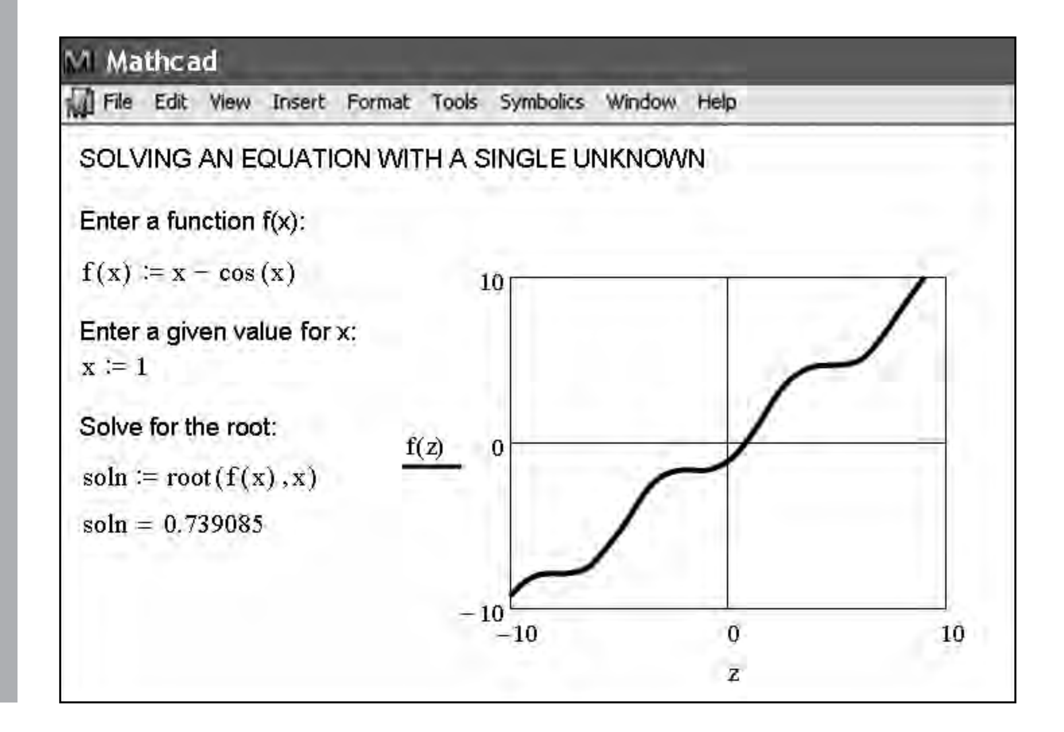


Figura 7.7
Pantalla Mathcad para encontrar la raíz de una sola ecuación.

un signo igual normal (=). El número de cifras significativas se ajusta mediante el menú desplegable de formato de número Format/Number. Las marcas de texto y las definiciones de ecuaciones se pueden colocar en cualquier parte de la pantalla con diversos tipos de letra, estilos, tamaños y colores. La gráfica se puede ubicar en cualquier parte de la hoja de trabajo haciendo clic en la ubicación deseada. Esto coloca una retícula roja en dicha ubicación. Luego use el menú de persiana Insert/Graph/X-Y Plot para colocar una gráfica vacía en la hoja de trabajo con espacios para las expresiones que se van a graficar y para los rangos de los ejes *x* y *y*. Simplemente teclee f(z) en el espacio sobre el eje *y*, y –10 y 10 para el rango del eje *z*. Mathcad hace todo el resto para producir la gráfica que se muestra en la figura 7.7. Una vez que se ha creado la gráfica, usted puede usar el menú de persiana Format/Graph/X-Y Plot para variar el tipo de gráfica; cambiar el color, el tipo y el grosor de la línea de trazo de la función; y agregue títulos, rótulos y otras características.

La figura 7.8 muestra cómo se puede usar Mathcad para encontrar las raíces de un polinomio usando la función **polyroots**. Primero, se introducen $p(x)$ y v usando := el símbolo de definición. Observe que v es un vector que contiene los coeficientes del polinomio comenzando con el término de grado cero y terminando en este caso con un término de tercer grado. La siguiente *r* se define (usando :=) como **polyroots(*v*)**, lo cual activa el método Laguerre. Las raíces contenidas en *r* se muestran como r^T usando un signo de igual normal (=). En seguida se construye una gráfica de manera similar a la anterior, salvo que ahora se usan dos variables de rango, *x* y *j*, para definir el rango del eje *x* y la ubicación de las raíces. La variable de rango para *x* se construye tecleando *x* y luego ":" (que aparece como :=) y luego –4, y después –3.99 y luego ";" (que Mathcad transforma en ..), y finalmente 4. Esto crea un vector de valor *x* con rango de –4 a 4, con un incremento de 0.01 para el eje *x*, con los valores correspondientes para $p(x)$ en el eje *y*. La variable de rango *j* se usa para crear tres valores para *r* y $p(r)$, que se grafican como círculos individuales pequeños. Observe que, nuevamente, en el esfuerzo de acomodar toda la sesión Mathcad en una sola pantalla, se ha colocado la gráfica encima de los comandos. Usted debería darse cuenta de que la gráfica tendría que estar debajo de los comandos para que funcione correctamente.

El último ejemplo muestra la solución de un sistema de ecuaciones no lineales usando un Mathcad Solve Block (figura 7.9). El proceso comienza usando el símbolo de definición para crear valores iniciales para *x* y *y*. La palabra **Given** alerta entonces a Mathcad de que lo que sigue es un sistema de ecuaciones. Luego siguen las ecuaciones y desigualdades (que no se usan aquí). Observe que, para esta aplicación, Mathcad necesita emplear un signo de igual simbólico que se teclea como

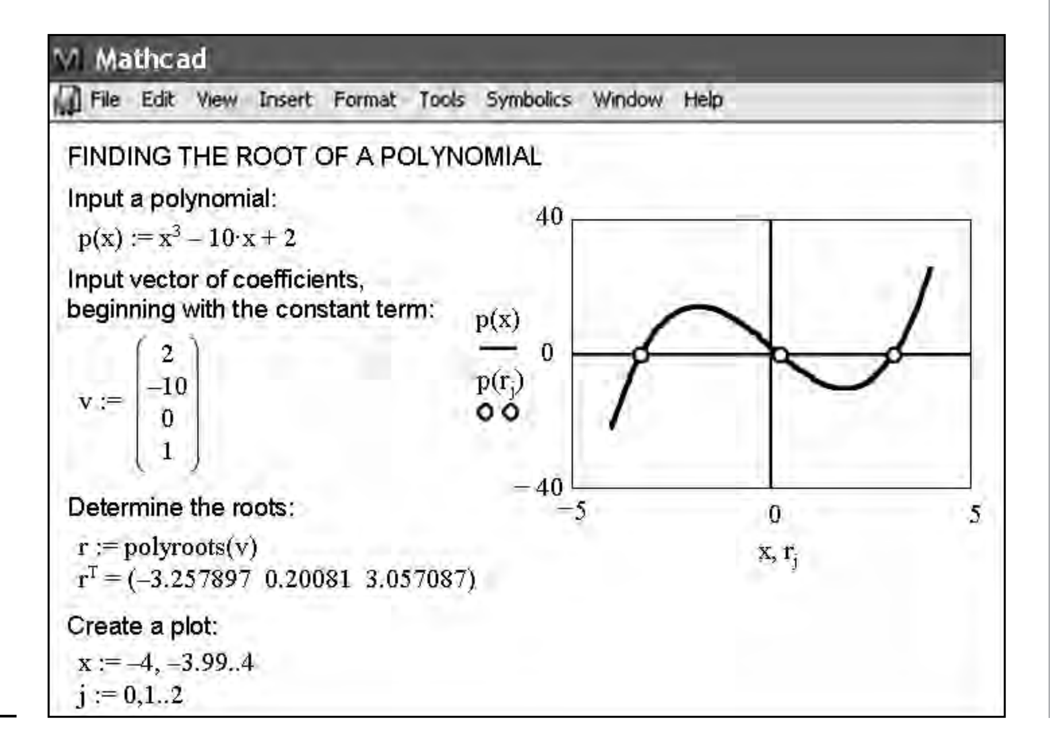


Figura 7.8
Pantalla Mathcad para despejar raíces de polinomio.

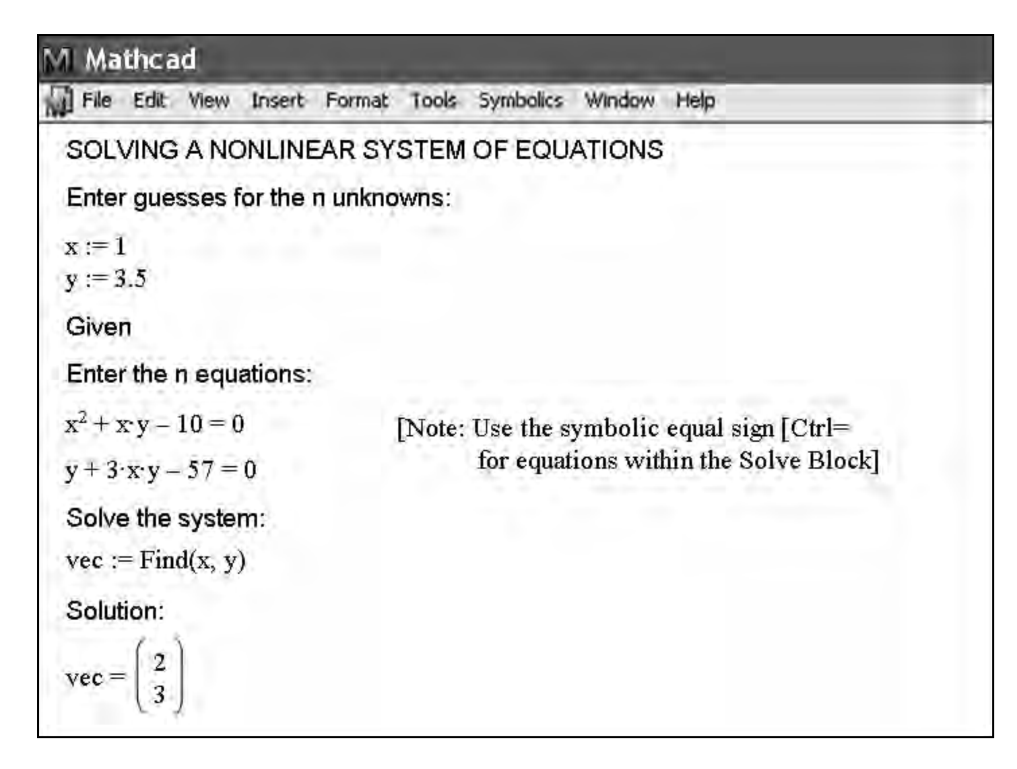


Figura 7.9
Pantalla Mathcad para resolver un sistema de ecuaciones no lineales.

[Ctrl]= o < y > para separar los lados izquierdo y derecho de una ecuación. Ahora, se define la variable vec como **Find** (x,y), y se muestra el valor de vec usando un signo de igual.

PROBLEMAS

7.1 Divida la ecuación polinómica $f(x) = x^4 - 7.5x^3 + 14.5x^2 + 3x - 20$ entre el factor monómico $x - 2$. ¿Es $x = 2$ una raíz?

7.2 Divida el polinomio $f(x) = x^5 - 5x^4 + x^3 - 6x^2 - 7x + 10$ entre el monomio factor $x - 2$.

7.3 Use el método de Müller para determinar la raíz real positiva de

a) $f(x) = x^3 + x^2 - 4x - 4$

b) $f(x) = x^3 - 0.5x^2 + 4x - 2$

7.4 Emplee el método de Müller o MATLAB para determinar las raíces reales y complejas de

a) $f(x) = x^3 - x^2 + 2x - 2$

b) $f(x) = 2x^4 + 6x^2 + 8$

c) $f(x) = x^4 - 2x^3 + 6x^2 - 2x + 5$

7.5 Use el método de Bairstow para determinar las raíces de

a) $f(x) = -2 + 6.2x - 4x^2 + 0.7x^3$

b) $f(x) = 9.34 - 21.97x + 16.3x^2 - 3.704x^3$

c) $f(x) = x^4 - 2x^3 + 6x^2 - 2x + 5$

7.6 Desarrolle un programa para implementar el método de Müller. Pruébelo con la repetición del ejemplo 7.2.

7.7 Emplee el programa que desarrolló en el problema 7.6 para determinar las raíces reales del problema 7.4*a*. Construya una gráfica (a mano, o con un paquete de software) para elegir valores iniciales apropiados.

7.8 Desarrolle un programa para implementar el método de Bairstow. Pruébelo con la repetición del ejemplo 7.3.

7.9 Use el programa que desarrolló en el problema 7.8 para determinar las raíces de las ecuaciones en el problema 7.5.

7.10 Determine la raíz real de $x^{3.5} = 80$ con Excel, MATLAB o Mathcad.

7.11 La velocidad de un paracaidista que cae está dada por

$$v = \frac{gm}{c}\left(1 - e^{-(c/m)t}\right)$$

donde $g = 9.81$ m/s^2. Para un paracaidista con un coeficiente de arrastre $c = 15$ kg/s, calcule la masa m de modo que la velocidad sea $v = 35$ m/s en $t = 8$ s. Use Excel, MATLAB o Mathcad para determinar *m*.

7.12 Determine las raíces de las ecuaciones no lineales simultáneas siguientes:

$$y = -x^2 + x + 0.75$$

$$y + 5xy = x^2$$

Utilice valores iniciales para $x = y = 1.2$ y use la herramienta Solver de Excel, o alguna librería o paquete que elija.

7.13 Determine las raíces de las ecuaciones no lineales simultáneas que siguen:

$$(x - 4)^2 + (y - 4)^2 = 5$$

$$x^2 + y^2 = 16$$

Use un método gráfico para obtener sus valores iniciales. Determine los estimados refinados con la herramienta Solver de Excel o alguna librería o paquete que elija.

7.14 Realice las operaciones MATLAB idénticas a las del ejemplo 7.7, o use algún software que elija, con objeto de encontrar todas las raíces de la ecuación polinomial

$$f(x) = (x + 2)(x + 5)(x - 6)(x - 4)(x - 8)$$

Obsérvese que es posible usar la función `poly` para convertir las raíces en un polinomio.

7.15 Use MATLAB o Mathcad para determinar las raíces de las ecuaciones en el problema 7.5.

7.16 Un cilindro circular de dos dimensiones se coloca en un flujo de velocidad alta y uniforme. Se desprenden vórtices del cilindro a frecuencia constante, la cual detectan sensores de presión en la superficie posterior del cilindro por medio de calcular qué tan seguido oscila la presión. Dados tres puntos de los datos, use el método de Müller para encontrar el momento en que la presión fue igual a cero.

Tiempo	0.60	0.62	0.64
Presión	20	50	60

7.17 Al tratar de encontrar la acidez de una solución de hidróxido de magnesio en ácido clorhídrico, se obtiene la ecuación siguiente:

$$A(x) = x^3 + 3.5x^2 - 40$$

donde x es la concentración del ion hidronio. Calcule la concentración del ion hidronio para una solución saturada (cuando la acidez es igual a cero) por medio de dos métodos diferentes en MATLAB (por ejemplo, en forma gráfica y la función `roots`).

7.18 Considere el sistema siguiente con tres incógnitas a, u y v:

$$u^2 - 2v^2 = a^2$$

$$u + v = 2$$

$$a^2 - 2a - u = 0$$

Encuentre los valores reales de las incógnitas, por medio de *a*) Solver de Excel, y *b*) algún paquete de software de manipulación simbólica.

7.19 En el análisis de sistemas de control, se desarrollan funciones de transferencia que relacionan en forma matemática la dinámica de la entrada de un sistema con su salida. La función de transferencia para un sistema de posicionamiento robotizado está dada por:

$$G(s) = \frac{C(s)}{N(s)} = \frac{s^3 + 9s^2 + 26s + 24}{s^4 + 5s^3 + 77s^2 + 153s + 90}$$

donde $G(s)$ = ganancia del sistema, $C(s)$ = salida del sistema, $N(s)$ = entrada del sistema y s = frecuencia compleja de la transformada de Laplace. Utilice una técnica numérica para obtener las raíces del numerador y el denominador, y factorícelas en la forma siguiente:

$$G(s) = \frac{(s + a_1)(s + a_2)(s + a_3)}{(s + b_1)(s + b_2)(s + b_3)(s + b_4)}$$

donde a_i y b_i = las raíces del numerador y el denominador, respectivamente.

7.20 Desarrolle una función de archivo M para el método de bisección, en forma similar a la de la figura 5.10. Pruebe la función por medio de repetir los cálculos de los ejemplos 5.3 y 5.4.

7.21 Desarrolle una función de archivo M para el método de la falsa posición. La estructura de su función debe ser similar al algoritmo de la bisección que se ilustra en la figura 5.10. Pruebe el programa por medio de repetir el ejemplo 5.5.

7.22 Desarrolle una función de archivo M para el método de Newton-Raphson, con base en la figura 6.4 y la sección 6.2.3. Junto con el valor inicial, introduzca como argumentos la función y derivada. Pruébelo con la repetición del cálculo del ejemplo 6.3.

7.23 Desarrolle una función de archivo M para el método de la secante, con base en la figura 6.4 y la sección 6.3.2. Junto con los dos valores iniciales, introduzca como argumento a la función. Pruébelo con la duplicación de los cálculos del ejemplo 6.6.

7.24 Desarrolle una función de archivo M para el método de la secante modificado, con base en la figura 6.4 y la sección 6.3.2. Junto con el valor inicial y la fracción de perturbación, introduzca como argumento a la función. Pruébelo con la duplicación de los cálculos del ejemplo 6.8.

CAPÍTULO 8

Estudio de casos: raíces de ecuaciones

La finalidad de este capítulo es utilizar los procedimientos numéricos analizados en los capítulos 5, 6 y 7 para resolver problemas de ingeniería reales. Las técnicas numéricas son importantes en aplicaciones prácticas, ya que con frecuencia los ingenieros encuentran problemas que no es posible resolver usando técnicas analíticas. Por ejemplo, modelos matemáticos simples que se pueden resolver analíticamente quizá no sean aplicables cuando se trata de problemas reales. Debido a esto, se deben utilizar modelos más complicados. En esta situación, es conveniente implementar una solución numérica en una computadora. En otros casos, los problemas de diseño en la ingeniería llegan a requerir soluciones de variables implícitas en ecuaciones complicadas.

Las siguientes aplicaciones son típicas de aquellas que en forma rutinaria se encuentran durante los últimos años de estudio y en estudios superiores. Más aún, son problemas representativos de aquellos que se encontrarán en la vida profesional. Los problemas provienen de las cuatro grandes ramas de la ingeniería: química, civil, eléctrica y mecánica. Dichas aplicaciones también sirven para ilustrar las ventajas y desventajas de las diversas técnicas numéricas.

La primera aplicación, tomada de la ingeniería química, proporciona un excelente ejemplo de cómo los métodos para determinar raíces permiten usar fórmulas realistas en la ingeniería práctica; además, demuestra de qué manera la eficiencia del método de Newton-Raphson se emplea cuando se requiere de un gran número de cálculos como método para la localización de raíces.

Los siguientes problemas de diseño en ingeniería se toman de las ingenierías civil, eléctrica y mecánica. La sección 8.2 usa la bisección para determinar cambios en la química del agua de lluvia debidos al aumento de dióxido de carbono en la atmósfera. En la sección 8.3 se explica cómo las raíces de ecuaciones trascendentes se usan en el diseño de un circuito eléctrico. En las secciones 8.2 y 8.3 también se muestra de qué forma los métodos gráficos ofrecen un conocimiento del proceso de localización de raíces. La sección 8.4 usa diversos métodos numéricos para calcular el factor de fricción para flujo de fluidos en un tubo.

8.1 LEYES DE LOS GASES IDEALES Y NO IDEALES (INGENIERÍA QUÍMICA Y BIOQUÍMICA)

Antecedentes La *ley de los gases ideales* está dada por

$$pV = nRT \tag{8.1}$$

donde p es la presión absoluta, V es el volumen, n es el número de moles, R es la constante universal de los gases y T es la temperatura absoluta. Aunque esta ecuación se utiliza ampliamente por los ingenieros y científicos, sólo es exacta en un rango limitado de presión y temperatura. Además, la ecuación (8.1) es apropiada solamente para algunos gases.

Una ecuación de estado alternativa para los gases está dada por:

$$\left(p + \frac{a}{v^2}\right)(v - b) = RT \tag{8.2}$$

conocida como la *ecuación de Van der Waals,* donde $v = V/n$ es el volumen molar, a y b son constantes empíricas que dependen del gas que se analiza.

Un proyecto de diseño en ingeniería química requiere que se calcule exactamente el volumen molar (v) del dióxido de carbono y del oxígeno para diferentes combinaciones de temperatura y

presión, de tal forma que los recipientes que contengan dichos gases se puedan seleccionar apropiadamente. También es importante examinar qué tan bien se apega cada gas a la ley de los gases ideales, comparando el volumen molar calculado con las ecuaciones (8.1) y (8.2). Se proporcionan los siguientes datos:

$R = 0.082054$ L atm/(mol K)

$$\left.\begin{aligned} a &= 3.592 \\ b &= 0.04267 \end{aligned}\right\} \text{dióxido de carbono}$$

$$\left.\begin{aligned} a &= 1.360 \\ b &= 0.03183 \end{aligned}\right\} \text{oxígeno}$$

Las presiones de diseño de interés son de 1, 10 y 100 atmósferas para combinaciones de temperatura de 300, 500 y 700 K.

Solución Los volúmenes molares de ambos gases se calculan usando la ley de los gases ideales, con $n = 1$. Por ejemplo, si $p = 1$ atm y $T = 300$ K,

$$v = \frac{V}{n} = \frac{RT}{p} = 0.082054 \frac{\text{L atm}}{\text{mol K}} \frac{300 \text{ K}}{1 \text{ atm}} = 24.6162 \text{ L/mol}$$

Estos cálculos se repiten para todas las combinaciones de presión y de temperatura que se presentan en la tabla 8.1.

Los cálculos del volumen molar a partir de la ecuación de Van der Waals se llevan a cabo usando cualquiera de los métodos numéricos para la determinación de raíces de ecuaciones analizados en los capítulos 5, 6 y 7, con

Tabla 8.1 Cálculos del volumen molar.

Temperatura, K	Presión, atm	Volumen molar (ley de los gases ideales), L/mol	Volumen molar (Van der Waals) Dióxido de carbono, L/mol	Volumen molar (Van der Waals) Oxígeno, L/mol
300	1	24.6162	24.5126	24.5928
	10	2.4616	2.3545	2.4384
	100	0.2462	0.0795	0.2264
500	1	41.0270	40.9821	41.0259
	10	4.1027	4.0578	4.1016
	100	0.4103	0.3663	0.4116
700	1	57.4378	57.4179	57.4460
	10	5.7438	5.7242	5.7521
	100	0.5744	0.5575	0.5842

$$f(v) = \left(p + \frac{a}{v^2}\right)(v - b) - RT \tag{8.3}$$

En este caso, como la derivada de $f(v)$ se determina fácilmente, entonces es conveniente y eficiente usar el método de Newton-Raphson. La derivada de $f(v)$ respecto a v está dada por

$$f'(v) = p - \frac{a}{v^2} + \frac{2ab}{v^3} \tag{8.4}$$

El método de Newton-Raphson se describe mediante la ecuación (6.6):

$$v_{i+1} = v_i - \frac{f(v_i)}{f'(v_i)}$$

la cual se utiliza para estimar la raíz. Por ejemplo, usando como valor inicial 24.6162, el volumen molar del dióxido de carbono a 300 K y 1 atmósfera es 24.5126 L/mol. Este resultado se obtuvo después de sólo dos iteraciones y tiene un ε_a menor del 0.001 por ciento.

En la tabla 8.1 se muestran resultados similares para todas las combinaciones de presión y de temperatura de ambos gases. Se observa que los resultados obtenidos con la ecuación de los gases ideales difieren de aquellos obtenidos usando la ecuación de Van der Waals, para ambos gases, dependiendo de los valores específicos de p y T. Además, como algunos de dichos resultados son significativamente diferentes, el diseño de los recipientes que contendrán a los gases podría ser muy diferente, dependiendo de qué ecuación de estado se haya empleado.

En este problema, se examinó una complicada ecuación de estado con el método de Newton-Raphson. En varios casos los resultados variaron de manera significativa respecto a la ley de los gases ideales. Desde un punto de vista práctico, el método de Newton-Raphson fue apropiado aquí, ya que $f'(v)$ resultó sencillo de calcular. De esta manera, es factible explotar las propiedades de rápida convergencia del método de Newton-Raphson.

Además de demostrar su poder en un solo cálculo, este problema de diseño muestra cómo el método de Newton-Raphson es especialmente atractivo cuando se requiere una gran cantidad de cálculos. Debido a la velocidad de las computadoras digitales, la eficiencia de varios métodos numéricos en la solución para la mayoría de las raíces de ecuaciones no se distingue en un cálculo único. Incluso una diferencia de 1 s entre el método de bisección y el eficiente método de Newton-Raphson no significa pérdida de tiempo cuando se realiza sólo un cálculo. Sin embargo, suponga que para resolver un problema se necesita calcular millones de raíces. En tal caso, la eficiencia del método podría ser un factor decisivo al elegir una técnica.

Por ejemplo, suponga que se requiere diseñar un sistema de control computarizado automático para un proceso de producción de sustancias químicas. Dicho sistema requiere una estimación exacta de volúmenes molares sobre una base esencialmente continua, para fabricar en forma conveniente el producto final. Se instalan medidores que proporcionan lecturas instantáneas de presión y temperatura. Se deben obtener valores de v para diversos gases que se usan en el proceso.

Para una aplicación como ésta, los métodos cerrados, como el de bisección o de la falsa posición, posiblemente consumirían mucho tiempo. Además, los dos valores iniciales que se requieren en estos métodos generarían un retraso crítico en el procedimiento. Dicho inconveniente afecta de igual forma al método de la secante, que también necesita dos valores iniciales.

En contraste, el método de Newton-Raphson requiere sólo de un valor inicial para determinar la raíz. La ley de los gases ideales podría emplearse para obtener un valor inicial del proceso. Después, suponiendo que el tiempo empleado sea lo bastante corto como para que la presión y la temperatura no varíen mucho entre los cálculos, la solución de la raíz anterior se puede usar como un buen valor inicial para la siguiente aplicación. De esta forma, se tendría de forma automática un valor inicial cercano a la solución, que es requisito indispensable para la convergencia del método de Newton-Raphson. Todas estas consideraciones favorecerán de buena manera la técnica de Newton-Raphson en estos problemas.

8.2 LOS GASES DE INVERNADERO Y LA LLUVIA (INGENIERÍA CIVIL Y AMBIENTAL)

Antecedentes La ingeniería civil es un amplio campo que abarca áreas tan diversas como las ingenierías estructural, geotérmica, de transporte, de recursos hidráulicos y ambiental. Esta última tradicionalmente se ha enfocado en el control de contaminación. Sin embargo, en años recientes, los ingenieros ambientales (así como los ingenieros químicos) se han centrado en problemas más amplios como el cambio climático.

Está bien documentado que los niveles atmosféricos de diversos *gases de invernadero* han ido aumentando durante los últimos 50 años. Por ejemplo, la figura 8.1 muestra datos para la presión parcial de dióxido de carbono (CO_2) recabados en Mauna Loa, Hawai, de 1958 a 2003. La tendencia de los datos se puede acomodar muy bien a un polinomio cuadrático (en la parte cinco aprenderá a determinar tales polinomios),

$$p_{CO_2} = 0.011825(t - 1\,980.5)^2 + 1.356975(t - 1\,980.5) + 339$$

donde p_{CO_2} = presión parcial del CO_2 en la atmósfera [ppm]. Los datos indican que los niveles han aumentado más de 19% durante el periodo, de 315 a 376 ppm.

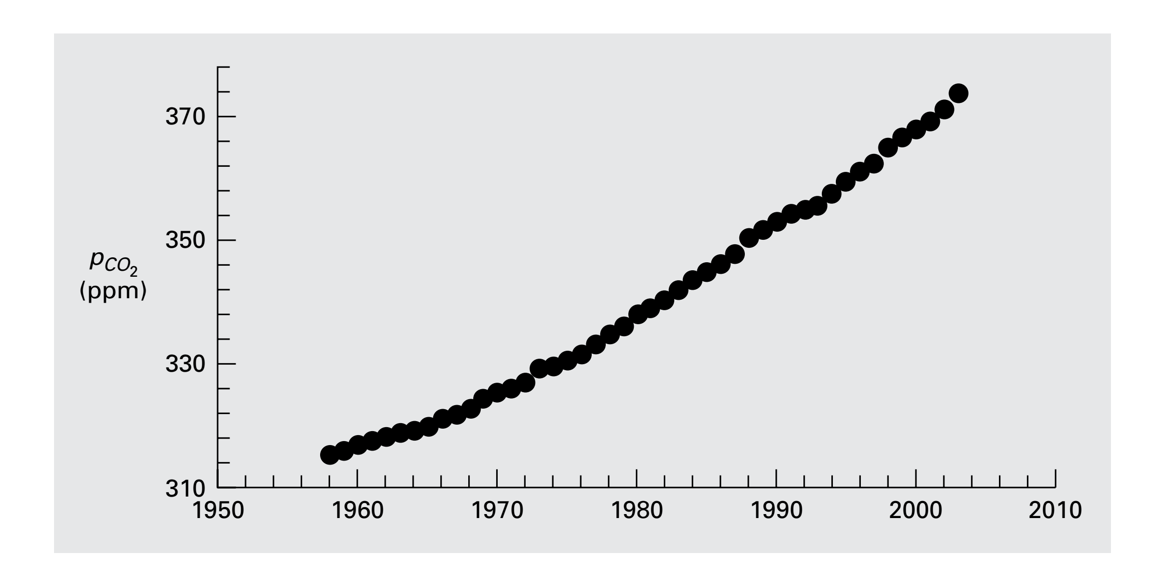


Figura 8.1
Promedio anual de presiones parciales de dióxido de carbono atmosférico [ppm] medidas en Mauna Loa, Hawai.

Además del calentamiento global, los gases de invernadero también pueden influir en la química atmosférica. Una pregunta que se puede hacer es cómo la tendencia de aumento en el dióxido de carbono está afectando el pH del agua de lluvia. Fuera de las zonas urbanas e industriales, está bien documentado que el dióxido de carbono es el principal determinante del pH de la lluvia. El pH es la medida de la actividad de los iones hidrógeno y, por lo tanto, de la acidez. Para soluciones acuosas diluidas, se puede calcular como

$$\text{pH} = -\log_{10}[\text{H}^+] \tag{8.5}$$

donde $[\text{H}^+]$ es la concentración molar de iones de hidrógeno.

El siguiente sistema no linear de ecuaciones determina la química del agua de lluvia:

$$K_1 = 10^6 \frac{[\text{H}^+][\text{HCO}_3^-]}{K_H\, p_{CO_2}} \tag{8.6}$$

$$K_2 = \frac{[\text{H}^+]\left[\text{CO}_3^{2-}\right]}{[\text{HCO}_3^-]} \tag{8.7}$$

$$K_w = [\text{H}^+][\text{OH}^-] \tag{8.8}$$

$$c_T = \frac{K_H\, p_{CO_2}}{10^6} + [\text{HCO}_3^-] + \left[\text{CO}_3^{2-}\right] \tag{8.9}$$

$$0 = [\text{HCO}_3^-] + 2\left[\text{CO}_3^{2-}\right] + [\text{OH}^-] - [\text{H}^+] \tag{8.10}$$

donde K_H = constante de Henry, y K_1, K_2 y K_w son los coeficientes de equilibrio. Las cinco incógnitas en el sistema de cinco ecuaciones no lineales son c_T = carbono inorgánico total, $[\text{HCO}_3^-]$ = bicarbonato, $[\text{CO}_3^{2-}]$ = carbonato, $[\text{H}^+]$ = ion de hidrógeno, $[\text{OH}^-]$ = ion de hidroxilo. Observe cómo aparece la presión parcial de CO_2 en las ecuaciones (8.6) y (8.9).

Use estas ecuaciones para calcular el pH del agua de lluvia dado que $K_H = 10^{-1.46}$, $K_1 = 10^{-6.3}$, $K_2 = 10^{-10.3}$ y $K_w = 10^{-14}$. Compare los resultados en 1958, cuando la p_{CO_2} era de 315, y en 2003, cuando era de 375 ppm. Al seleccionar un método numérico para su cálculo, considere lo siguiente:

- Usted sabe con certeza que el pH de la lluvia en áreas limpias siempre cae entre 2 y 12.
- También sabe que sus dispositivos de medición sólo pueden medir pH con una precisión de hasta dos cifras decimales.

Solución Hay diversas maneras de resolver este sistema no lineal de cinco ecuaciones. Una forma es eliminar incógnitas combinándolas para producir una sola función que sólo depende de $[H^+]$. Para hacer esto, primero resuelva las ecuaciones (8.6) y (8.7) para despejar:

$$[HCO_3^-] = \frac{K_1}{10^6[H^+]} K_H p_{CO_2} \tag{8.11}$$

$$[CO_3^{2-}] = \frac{K_2[HCO_3^-]}{[H^+]} \tag{8.12}$$

Sustituya la ecuación (8.11) en la (8.12)

$$\left[CO_3^{2-}\right] = \frac{K_2 K_1}{10^6[H^+]^2} K_H p_{CO_2} \tag{8.13}$$

Las ecuaciones (8.11) y (8.13) se pueden sustituir junto con la ecuación (8.8) en la ecuación (8.10) para dar

$$0 = \frac{K_1}{10^6[H^+]} K_H p_{CO_2} + 2\frac{K_2 K_1}{10^6[H^+]^2} K_H p_{CO_2} + \frac{K_w}{[H^+]} - [H^+] \tag{8.14}$$

Aunque podría no ser evidente, este resultado es un polinomio de tercer grado en $[H^+]$. Así, su raíz se puede usar para calcular el pH del agua de lluvia.

Ahora debemos decidir qué método numérico debemos emplear para obtener la solución. Hay dos razones por las que la bisección sería una buena elección. La primera es que el hecho de que el pH siempre cae dentro del rango de 2 a 12 nos proporciona dos valores iniciales; la segunda razón es que, como el pH se puede medir sólo hasta dos cifras decimales de precisión, será satisfactorio un error absoluto de $E_{a,d} = 0.005$. Recuerde que, dado un acontecimiento inicial y el error relativo deseado, podemos calcular *a priori* el número de iteraciones. Usando la ecuación (5.5), el resultado es $n = \log_2(10)/0.005 = 10.9658$. Así, once iteraciones de bisección producirán la precisión deseada.

Si se hace esto, el resultado para 1958 será un pH de 5.6279, con un error relativo de 0.0868%. Se puede confiar en que el resultado redondeado de 5.63 es correcto hasta 2 cifras decimales. Esto se puede verificar realizando otra corrida con más iteraciones. Por ejemplo, si se realizan 35 iteraciones, se obtiene un resultado de 5.6304, con un error relativo aproximado de $\varepsilon_a = 5.17 \times 10^{-9}\%$. Se puede repetir el mismo cálculo para las condiciones de 2003 para dar pH = 5.59, con ε_a = 0.0874%.

Es interesante que estos resultados indican que una elevación de 19% en los niveles de CO_2 atmosférico ha producido sólo una caída de 0.67% en el pH. Aunque esto es ciertamente verdadero, recuerde que el pH representa una escala logarítmica como se define en la ecuación (8.5). En consecuencia, una caída de 1 unidad en pH representa un aumento de 10 veces en iones de hidrógeno. La concentración se puede calcular como $[H^+] = 10^{-pH}$, y el cambio resultante en el porcentaje se puede calcular como 9.1%. Por lo tanto, la concentración de iones de hidrógeno ha aumentado en alrededor de 9%.

Hay una considerable controversia en relación con la verdadera importancia de las tendencias de los gases de invernadero. Sin embargo, independientemente de las implicaciones finales, es bastante aleccionador darse cuenta de que algo tan grande como nuestra atmósfera ha cambiado tanto a lo largo de un periodo relativamente corto. Este estudio de caso ilustra cómo se pueden emplear los métodos numéricos para analizar e interpretar tales tendencias. Durante los próximos años se espera que los ingenieros y los científicos puedan usar estas herramientas para lograr un mejor entendimiento y ayudar a la racionalización del debate sobre sus ramificaciones.

8.3 DISEÑO DE UN CIRCUITO ELÉCTRICO (INGENIERÍA ELÉCTRICA)

Antecedentes Los ingenieros eléctricos emplean las leyes de Kirchhoff para estudiar el comportamiento de los circuitos eléctricos en estado estacionario (que no varía con el tiempo). En la sección 12.3 se analiza el comportamiento de dichos estados estacionarios. Otro problema importante tiene que ver con circuitos de naturaleza transitoria, donde súbitamente ocurren cambios temporales. Esta situación se presenta cuando se cierra el interruptor como en la figura 8.2. En tal caso, existe un periodo de ajuste al cerrar el interruptor hasta que se alcance un nuevo estado estacionario. La longitud de este periodo de ajuste está íntimamente relacionada con las propiedades de almacenamiento de energía, tanto del capacitor como del inductor. La energía almacenada puede oscilar entre estos dos elementos durante un periodo transitorio. Sin embargo, la resistencia en el circuito disipará la magnitud de las oscilaciones.

El flujo de corriente a través del resistor provoca una caída de voltaje (V_R), dada por

$$V_R = iR$$

donde i = la corriente y R = la resistencia del resistor. Si las unidades de R e i son ohms y amperes, respectivamente, entonces las unidades de V_R son voltios.

De manera semejante, un inductor se opone a cambios de corriente tales que la caída del voltaje a través del inductor V_L es

$$V_L = L\frac{di}{dt}$$

donde L = la inductancia. Si las unidades de L e i son henrios y amperes, respectivamente, entonces las de V_L son voltios, y las de t son segundos.

La caída del voltaje a través del capacitor (V_C) depende de la carga (q) sobre éste:

$$V_C = \frac{q}{C} \tag{8.15}$$

Figura 8.2
Un circuito eléctrico. Cuando se cierra el interruptor, la corriente experimenta una serie de oscilaciones hasta que se alcance un nuevo estado estacionario.

donde C = la capacitancia. Si las unidades de carga se expresan en coulombios, entonces la unidad de C es el faradio.

La segunda ley de Kirchhoff establece que la suma algebraica de las caídas de voltaje alrededor de un circuito cerrado es cero. Así que, después de cerrar el interruptor se tiene

$$L\frac{di}{dt} + Ri + \frac{q}{C} = 0 \tag{8.16}$$

Sin embargo, como la corriente se relaciona con la carga de acuerdo con

$$i = \frac{dq}{dt} \tag{8.17}$$

Por lo tanto,

$$L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{1}{C}q = 0 \tag{8.18}$$

Ésta es una ecuación diferencial ordinaria lineal de segundo grado que se resuelve usando los métodos de cálculo. Esta solución está dada por

$$q(t) = q_0 e^{-Rt/(2L)} \cos\left[\sqrt{\frac{1}{LC} - \left(\frac{R}{2L}\right)^2}\, t\right] \tag{8.19}$$

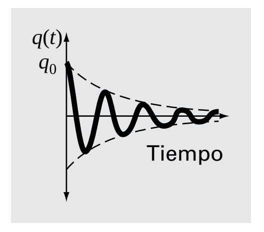


Figura 8.3
La carga en un capacitor como función del tiempo después de cerrar el interruptor de la figura 8.2.

donde $t = 0$, $q = q_0 = V_0C$ y V_0 = el voltaje de la batería. La ecuación (8.19) describe la variación de la carga en el capacitor. La solución $q(t)$ se grafica en la figura 8.3.

Un problema de diseño típico en ingeniería eléctrica consistiría en la determinación del resistor apropiado para disipar energía a una razón especificada, con valores conocidos de L y C. En este problema, suponga que la carga se debe disipar a 1% de su valor original ($q/q_0 = 0.01$) en $t = 0.05$ s, con $L = 5$ H y $C = 10^{-4}$F.

Solución Es necesario despejar R de la ecuación (8.19) con valores conocidos para q, q_0, L y C. Sin embargo, debe emplear una técnica de aproximación numérica, ya que R es una variable implícita en la ecuación (8.19). Se usará el método de bisección para dicho propósito. Los otros métodos estudiados en los capítulos 5 y 6 también son apropiados; aunque el método de Newton-Raphson tiene el inconveniente de que la derivada de la ecuación (8.19) es un poco complicada. Reordenando la ecuación (8.19),

$$f(R) = e^{-Rt/(2L)} \cos\left[\sqrt{\frac{1}{LC} - \left(\frac{R}{2L}\right)^2}\, t\right] - \frac{q}{q_0}$$

utilizando los valores numéricos dados,

$$f(R) = e^{-0.005R} \cos\left[\sqrt{2\,000 - 0.01R^2}\,(0.05)\right] - 0.01 \tag{8.20}$$

Examinando esta ecuación sugiere que un rango inicial razonable para R es 0 a 400 Ω (ya que $2\,000 - 0.01R^2$ debe ser mayor que cero). La figura 8.4 es una gráfica de la ecuación (8.20), que confirma lo anterior. Al hacer veintiún iteraciones con el método de bisección se obtiene una raíz aproximada $R = 328.1515$ Ω, con un error menor al 0.0001 por ciento.

De esta forma, se especifica un resistor con este valor para el circuito mostrado en la figura 8.2 y se espera tener una disipación consistente con los requisitos del problema. Este problema de diseño no se podría resolver eficientemente sin el uso de los métodos numéricos vistos en los capítulos 5 y 6.

8.4 FRICCIÓN EN TUBOS (INGENIERÍA MECÁNICA Y AEROESPACIAL)

Antecedentes La determinación del flujo de fluidos en tuberías y mangueras tiene gran importancia en muchas áreas de la ingeniería y de la ciencia. En la ingeniería mecánica y en la aeroespacial, las aplicaciones típicas incluyen el flujo de líquidos y gases a través de sistemas de enfriamiento.

Los parámetros de la resistencia al flujo en tales conductos están determinados por un número adimensional que se llama *factor de fricción*. Para flujo turbulento, la *ecuación de Colebrook* proporciona un medio para calcular el factor de fricción,

$$0 = \frac{1}{\sqrt{f}} + 2.0 \log\left(\frac{\varepsilon}{3.7D} + \frac{2.51}{\text{Re}\sqrt{f}}\right) \tag{8.21}$$

donde ε = rugosidad (m), D = diámetro (m) y Re = el *número de Reynolds*,

$$\text{Re} = \frac{\rho VD}{\mu}$$

donde ρ = densidad del fluido [kg/m^3], V = velocidad [m/s] y μ = viscosidad dinámica [N·s/m^2]. Además de aparecer en la ecuación (8.21), el número de Reynolds también sirve como criterio de si el flujo es turbulento (Re > 4 000).

En el presente estudio de caso, se ilustrará cómo se pueden utilizar los métodos numéricos que se tratan en esta parte del libro para determinar f para flujo de aire por un tubo liso y delgado.

Para este caso, los parámetros son $\rho = 1.23$ kg/m^3, $\mu = 1.79 \times 10^{-5}$ N · s/m^2, $D = 0.005$ m, $V = 40$ m/s y $\varepsilon = 0.0015$ mm. Observe que los factores de fricción están comprendidos entre alrededor de 0.008 y 0.08. Además, la formulación explícita, llamada *ecuación de Swamee-Jain*, proporciona un cálculo aproximado,

$$f = \frac{1.325}{\left[\ln\left(\frac{\varepsilon}{3.7D} + \frac{5.74}{\text{Re}^{0.9}}\right)\right]^2} \tag{8.22}$$

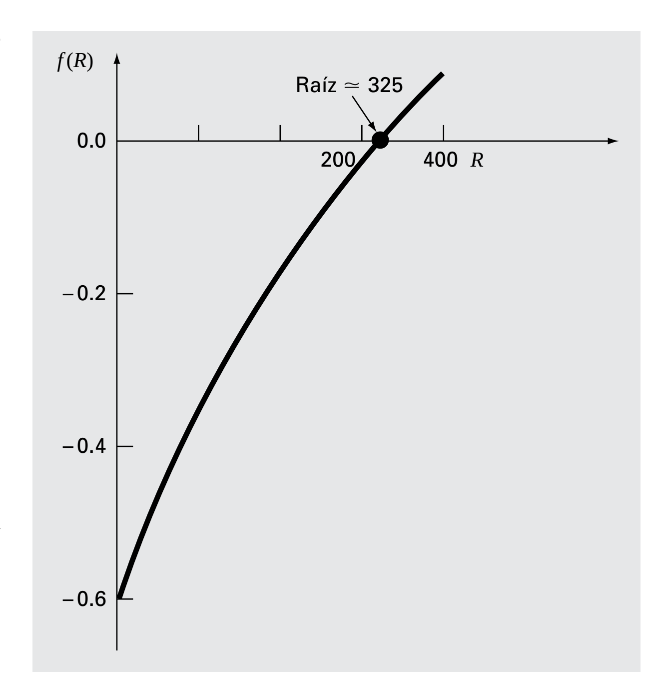


Figura 8.4
Gráfica de la ecuación (8.20) que se usa para obtener los valores iniciales de R que abarcan la raíz.

Solución El número de Reynolds se puede calcular como

$$\text{Re} = \frac{\rho VD}{\mu} = \frac{1.23(40)0.005}{1.79 \times 10^{-5}} = 13\,743$$

Este valor, junto con los otros parámetros, se puede sustituir en la ecuación (8.21) para dar

$$g(f) = \frac{1}{\sqrt{f}} + 2.0\log\left(\frac{0.0000015}{3.7(0.005)} + \frac{2.51}{13\,743\sqrt{f}}\right)$$

Antes de determinar la raíz, es aconsejable graficar la función para estimar los valores iniciales y para anticipar dificultades posibles. Esto se puede hacer fácilmente con herramientas como MATLAB, Excel o Mathcad. Por ejemplo, es posible generar una gráfica de la función con los siguientes comandos de MATLAB

```
>> rho=1.23;mu=1.79e-5;D=0.005;V=40;e=0.0015/1000;
>> Re=rho*V*D/mu;
>> g=@(f) 1/sqrt(f)+2*log10(e/(3.7*D)+2.51/(Re*sqrt(f)));
>> fplot(g,[0.008 0.08]),grid,xlabel('f'),ylabel('g(f)')
```

Como en la figura 8.5, la raíz se ubica en alrededor de 0.03.

Como se dan los valores iniciales ($x_l = 0.008$ y $x_u = 0.08$), se puede usar cualquiera de los métodos cerrados del capítulo 5. Por ejemplo, la bisección da un valor de $f = 0.0289678$, con un error porcentual relativo del error de 5.926×10^{-5} en 22 iteraciones. El método de falsa posición da un resultado de precisión similar en 26 iteraciones. Así, aunque producen el resultado correcto, son algo ineficientes. Esto no sería importante para una sola aplicación, pero podría resultar prohibitivo si se hicieran muchas evaluaciones.

Se podría tratar de obtener un desempeño mejorado recurriendo a un método abierto. Como la ecuación (8.21) es relativamente fácil de diferenciar, el método de Newton-Raphson es un buen candidato. Por ejemplo, usando un valor inicial en el extremo inferior del rango ($x_0 = 0.008$). Newton-Raphson converge rápidamente a 0.0289678, con un error aproximado de 6.87×10^{-6}% en sólo 6 iteraciones. Sin embargo, cuando el valor inicial se ubica en el extremo superior del rango ($x_0 = 0.08$), ¡la rutina diverge!

Como se puede ver al observar la figura 8.5, esto sucede porque la pendiente de la función en el valor inicial hace que la primera iteración salte a un valor negativo. Las corridas siguientes demuestran que, para este caso, la convergencia sólo ocurre cuando el valor inicial es menor de alrededor de 0.066.

De modo que se puede ver que, aunque Newton-Raphson es muy eficiente, necesita buenos valores iniciales. Para la ecuación de Colebrook, una buena estrategia sería emplear la ecuación de Swamee-Jain [ecuación (8.22)] para dar un valor inicial como en

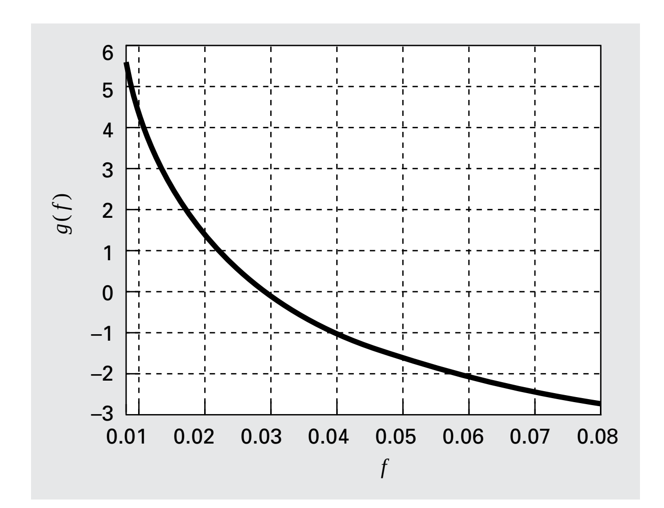


Figura 8.5

$$f = \frac{1.325}{\left[\ln\left(\frac{0.0000015}{3.7(0.005)} + \frac{5.74}{13743^{0.9}}\right)\right]^2} = 0.029031$$

Para este caso, Newton-Raphson converge en sólo 3 iteraciones rápidamente a 0.0289678 con un error aproximado de 8.51×10^{-10}%.

Además de estas funciones simples, se pueden también usar determinadores profesionales de raíces, como la función incorporado `fzero` de MATLAB. Sin embargo, al igual que con el método Newton-Raphson, también ocurre divergencia cuando se usa la función `fzero` con un solo valor. Sin embargo, en este caso, los valores en el extremo inferior del rango causan problemas. Por ejemplo,

```
>> rho=1.23;mu=1.79e-5;D=0.005;V=40;e=0.0015/1000;
>> Re=rho*V*D/mu
>> g=@(f) 1/sqrt(f)+2*log10(e/(3.7*D)+2.51/(Re*sqrt(f)));
>> fzero(g,0.008)

Exiting fzero: aborting search for an interval containing a
sign change because complex function value encountered
during search. (Function value at -0.0028 is -4.92028-
20.2423i.)
Check function or try again with a different starting value.
ans =
   NaN
```

Si las iteraciones se muestran en pantalla usando `optimset` (recuerde la sección 7.7.2), se revela que ocurre un valor negativo durante la fase de búsqueda antes de que se detecte el cambio de signo y la rutina aborte. Sin embargo, para valores iniciales únicos superiores de alrededor de 0.016, la rutina trabaja muy bien. Por ejemplo, para el valor de 0.08 que causó problemas para el método Newton-Raphson, `fzero` funciona muy bien.

```
>> fzero(g,0.08)

ans =
   0.02896781017144
```

Y, como nota final, veamos si es posible la convergencia para iteración sencilla de punto fijo. La versión más fácil y directa consiste en despejar la primera f en la ecuación (8.21),

$$f_{i+1} = \frac{0.25}{\left(\log\left(\frac{\varepsilon}{3.7D} + \frac{2.51}{\text{Re}\sqrt{f_i}}\right)\right)^2} \tag{8.23}$$

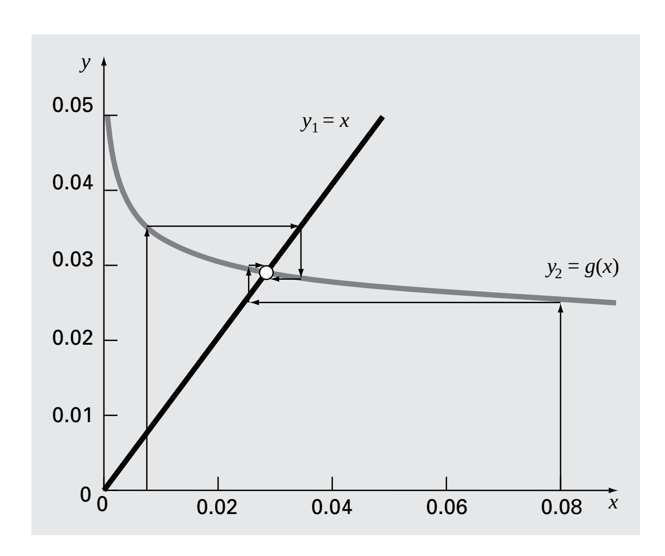


Figura 8.6

La gráfica de 2 curvas de esta función indica un resultado sorprendente (figura 8.6). Recuerde que la iteración de punto fijo converge cuando la curva y_2 tiene una pendiente relativamente baja (es decir, $|g'(\xi)| < 1$). Como se indica en la figura 8.6, el hecho de que la curva y_2 tenga pendiente bastante baja en el rango de f = 0.008 a 0.08, significa que la iteración de punto fijo no sólo converge, sino que lo hace con bastante rapidez. De hecho, para valores iniciales entre 0.008 y 0.08, la iteración de punto fijo da predicciones con errores relativos porcentuales menores de 0.008% en seis o menos iteraciones. Así, este sencillo método que sólo necesita un valor y no necesita estimados de derivada funciona bastante bien para este caso particular.

El mensaje que se debe sacar de este estudio de caso es que incluso un gran software, profesionalmente desarrollado como el MATLAB no siempre es a prueba de errores. Además, normal-

mente no hay un solo método que funcione mejor para los problemas. Los usuarios experimentados entienden las fuerzas y las debilidades de las técnicas numéricas disponibles. Además, entienden suficiente de la teoría subyacente, de modo que pueden manejar eficazmente situaciones en las que un método falla.

PROBLEMAS

Ingeniería química/ingeniería bioquímica

8.1 Realice el mismo cálculo que en la sección 8.1, pero para acetona (a = 12.02 y b = 0.08407) a una temperatura de 375 K y p de 2.0 atm. Compare sus resultados con la ley de gases ideales. Use cualquiera de los métodos numéricos explicados en los capítulos 5 y 6 para realizar el cálculo. Justifique su elección de técnica.

8.2 En ingeniería química, los reactores de flujo tipo tapón (es decir, aquellos en que el fluido va de un extremo al otro con una mezcla mínima a lo largo del eje longitudinal) se usan para convertir reactantes en productos. Se ha determinado que la eficiencia de la conversión algunas veces se mejora recirculando una porción de la corriente del producto, de tal forma que regrese a la entrada para un paso adicional a través del reactor (figura P8.2). La razón de recirculación se define como

$$R = \frac{\text{volumen de fluido que regresa a la entrada}}{\text{volumen que sale del sistema}}$$

Suponga que se está procesando una sustancia química A para generar un producto B. Para el caso en que A forma a B de acuerdo con una reacción autocatalítica (es decir, en la cual uno de los productos actúa como catalizador o estimulante en la reacción), es posible demostrar que una razón óptima de recirculación debe satisfacer

$$\ln\frac{1+R(1-X_{Af})}{R(1-X_{Af})} = \frac{R+1}{R[1+R(1-X_{Af})]}$$

donde X_{Af} es la fracción del reactante A que se convierte en el producto B. La razón óptima de recirculación corresponde a un reactor de tamaño mínimo necesario para alcanzar el nivel deseado de conversión. Utilice un método numérico para determinar la razón de recirculación necesaria, de manera que se minimice el tamaño del reactor para una conversión fraccional de X_{Af} = 0.9.

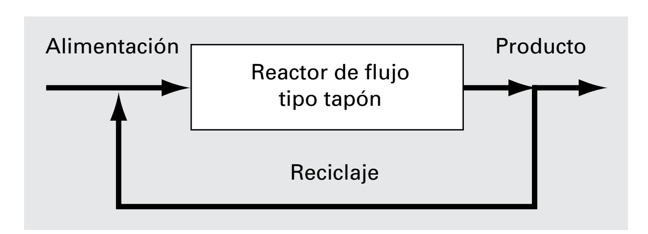


Figura P8.2
Representación esquemática de un reactor de flujo tipo tapón con recirculación.

8.3 En un proceso de ingeniería química el vapor de agua (H_2O) se calienta a temperaturas lo suficientemente altas para que una porción significativa del agua se disocie, o se rompa, para formar oxígeno (O_2) e hidrógeno (H_2):

$$H_2O \rightleftharpoons H_2 + \tfrac{1}{2}O_2$$

Si se supone que ésta es la única reacción que se lleva a cabo, la fracción molar x de H_2O que se disocia se representa por

$$K = \frac{x}{1-x}\sqrt{\frac{2p_t}{2+x}} \qquad \text{(P8.3.1)}$$

donde K = la constante de equilibrio de la reacción y p_t = la presión total de la mezcla. Si p_t = 3 atm y K = 0.05, determine el valor de x que satisfaga la ecuación (P8.3.1).

8.4 La siguiente ecuación permite calcular la concentración de un químico en un reactor donde se tiene una mezcla completa:

$$c = c_{\text{ent}}(1 - e^{-0.04t}) + c_0 e^{-0.04t}$$

Si la concentración inicial es c_0 = 4 y la concentración de entrada es c_{ent} = 10, calcule el tiempo requerido para que c sea el 93% de c_{ent}.

8.5 Una reacción química reversible

$$2A + B \rightleftharpoons C$$

se puede caracterizar por su relación de equilibrio

$$K = \frac{c_c}{c_a^2 c_b}$$

donde la nomenclatura c_i representa la concentración del componente i. Suponga que definimos una variable x para representar el número de moles de C que se producen. Se puede usar la conservación de la masa para reformular la relación de equilibrio como

$$K = \frac{(c_{c,0} + x)}{(c_{a,0} - 2x)^2(c_{b,0} - x)}$$

donde el subíndice 0 designa la concentración inicial de cada componente. Si K = 0.015, $c_{a,0}$ = 42, $c_{b,0}$ = 30 y $c_{c,0}$ = 4, determine el valor de x. *a*) Obtenga la solución gráficamente. *b*) Con base en *a*), despeje la raíz con valores iniciales de x_l = 0 y x_u = 20 hasta ε_s = 0.5%. Elija ya sea bisección o falsa posición para obtener su solución. Justifique su elección.

8.6 Las siguientes reacciones químicas se llevan a cabo en un sistema cerrado

$$2A + B \rightleftharpoons C$$
$$A + D \rightleftharpoons C$$

En equilibrio, éstas pueden caracterizarse por

$$K_1 = \frac{c_c}{c_a^2 c_b}$$
$$K_2 = \frac{c_c}{c_a c_d}$$

donde la nomenclatura representa la concentración del componente i. Si x_1 y x_2 son el número de moles de C que se producen debido a la primera y segunda reacciones, respectivamente, emplee un método similar al del problema 8.5 para reformular las relaciones de equilibrio en términos de las concentraciones iniciales de los componentes. Después, use el método de Newton-Raphson para resolver el par

de ecuaciones simultáneas no lineales para x_1 y x_2 si $K_1 = 4 \times 10^{-4}$, $K_2 = 3.7 \times 10^{-2}$, $c_{a,0} = 50$, $c_{b,0} = 20$, $c_{c,0} = 5$ y $c_{d,0} = 10$. Utilice un método gráfico para proponer los valores iniciales.

8.7 La ecuación de estado de Redlich-Kwong está dada por

$$p = \frac{RT}{v - b} - \frac{a}{v(v + b)\sqrt{T}}$$

donde R = la constante universal de los gases [= 0.518 kJ/(kg K)], T = temperatura absoluta (K), p = presión absoluta (kPa) y v = volumen de un kg de gas (m^3/kg). Los parámetros a y b se calculan mediante

$$a = 0.427\frac{R^2T_c^{2.5}}{p_c} \qquad b = 0.0866R\frac{T_c}{p_c}$$

donde p_c = presión crítica (kPa) y T_c = temperatura crítica (K). Como ingeniero químico, se le pide determinar la cantidad de combustible metano (p_c = 4 600 kPa y T_c = 191 K) que se puede almacenar en un tanque de 3 m^3 a una temperatura de –40°C con una presión de 65 000 kPa. Emplee el método de localización de raíces de su elección para calcular v y luego determine la masa de metano contenida en el tanque.

8.8 El volumen V de un líquido contenido en un tanque horizontal cilíndrico de radio r y longitud L está relacionado con la profundidad del líquido h por

$$V = \left[r^2\cos^{-1}\left(\frac{r - h}{r}\right) - (r - h)\sqrt{2rh - h^2}\right]L$$

Determine h dados r = 2 m, L = 5 m y V = 8 m^3. Observe que si usted está usando un lenguaje de programación o una herramienta de software que no tenga abundancia de funciones trigonométricas, el arco coseno se puede calcular con

$$\cos^{-1}x = \frac{\pi}{2} - \tan^{-1}\left(\frac{x}{\sqrt{1 - x^2}}\right)$$

8.9 El volumen V del líquido contenido en un tanque esférico de radio r está relacionado con la profundidad h del líquido por

$$V = \frac{\pi h^2(3r - h)}{3}$$

Determine h para r = 1 m y V = 0.5 m^3.

8.10 Para el tanque esférico del problema 8.9, es posible desarrollar las siguientes dos fórmulas para el método de punto fijo:

$$h = \sqrt{\frac{h^3 + (3V/\pi)}{3r}}$$

y

$$h = \sqrt[3]{3\left(rh^2 - \frac{V}{\pi}\right)}$$

Si r = 1 m y V = 0.5 m^3, determine si cualquiera de las dos alturas es estable, y el rango de valores iniciales para los que sí son estables.

8.11 La operación de un reactor de flujo tipo tapón de densidad constante para la producción de una sustancia por una reacción enzimática se describe mediante la ecuación que se muestra en seguida, donde V es el volumen del reactor, F es el caudal del reactante C, C_{ent} y C_{sal} son las concentraciones del reactante a la entrada y a la salida del reactor, respectivamente, y K y $k_{máx}$ son constantes. Para un reactor de 100 L, con una concentración de entrada de C_{ent} = 0.2 M, un caudal de entrada de 80 L/s, $k_{máx} = 10^{-2}$ s^{-1} y K = 0.1 M, encuentre la concentración de C a la salida del reactor.

$$\frac{V}{F} = -\int_{C_{ent}}^{C_{sal}} \frac{K}{k_{máx}C} + \frac{1}{k_{máx}}dC$$

8.12 *La ecuación de Ergun*, que se da más adelante, sirve para describir el flujo de un líquido a través de un cauce comprimido. ΔP es la caída de presión, ρ es la densidad del fluido, G_o es la velocidad másica (el cociente del flujo de masa dividido entre el área de la sección transversal), D_p es el diámetro de las partículas dentro del cauce, μ es la viscocidad del fluido, L es la longitud del cauce y ε es la fracción vacía del cauce.

$$\frac{\Delta P\rho}{G_o^2}\frac{D_p}{L}\frac{\varepsilon^3}{1 - \varepsilon} = 150\frac{1 - \varepsilon}{\left(\dfrac{D_pG_o}{\mu}\right)} + 1.75$$

Dados los siguientes valores para los parámetros encuentre la fracción vacía ε del cauce.

$$\frac{D_pG_o}{\mu} = 1\,000$$

$$\frac{\Delta P\rho D_p}{G_o^2L} = 10$$

8.13 En una sección de tubo, la caída de presión se calcula así:

$$\Delta p = f\frac{L\rho V^2}{2D}$$

donde Δp = caída de presión (Pa), f = factor de fricción, L = longitud del tubo [m], ρ = densidad (kg/m^3), V = velocidad (m/s) y D = diámetro (m). Para el flujo turbulento, la *ecuación de Colebrook* proporciona un medio para calcular el factor de fricción,

$$\frac{1}{\sqrt{f}} = -2.0\log\left(\frac{\varepsilon}{3.7D} + \frac{2.51}{\text{Re}\sqrt{f}}\right)$$

donde ε = rugosidad (m) y Re = *número de Reynolds*,

$$\text{Re} = \frac{\rho VD}{\mu}$$

donde μ = viscosidad dinámica (N · s/m^2).

a) Determine Δp para un tramo horizontal de tubo liso de 0.2 m de longitud, dadas ρ = 1.23 kg/m^3, $\mu = 1.79 \times 10^{-5}$ N · s/m^2, D = 0.005 m, V = 40 m/s y ε = 0.0015 mm. Utilice un método numérico para determinar el factor de fricción. Obsérvese que los tubos lisos tienen Re < 10^5, un valor inicial apropiado se obtiene con el uso de la *fórmula de Blasius*, $f = 0.316/\text{Re}^{0.25}$.

b) Repita el cálculo pero para un tubo de acero comercial más rugoso (ε = 0.045 mm).

Ingeniería civil y ambiental

8.14 En ingeniería estructural, la *fórmula de la secante* define la fuerza por unidad de área, P/A, que ocasiona la tensión máxima σ_m en una columna que tiene una razón de esbeltez L/k dada es:

$$\frac{P}{A} = \frac{\sigma_m}{1 + (ec/k^2)\sec\,[0.5\sqrt{P/(EA)}(L/k)]}$$

donde ec/k^2 = razón de excentricidad y E = módulo de elasticidad. Si para una viga de acero, E = 200 000 MPa, ec/k^2 = 0.2 y σ_m = 250 MPa, calcule P/A para L/k = 100. Recuerde que sec x = 1/cos x.

8.15 En la ingeniería ambiental (una especialidad de la ingeniería civil), la ecuación siguiente se emplea para calcular el nivel de oxígeno c (mg/L) en un río aguas abajo de la descarga de un drenaje:

$$c = 10 - 20(e^{-0.2x} - e^{-0.75x})$$

donde x es la distancia aguas abajo en kilómetros.

a) Determine la distancia aguas abajo de la corriente, a la cual el nivel de oxígeno cae hasta una lectura de 5 mg/L. (*Sugerencia:* Está dentro de 2 km de la descarga.) Encuentre la respuesta con un error de 1%. Obsérvese que los niveles de oxígeno por debajo de 5 mg/L por lo general son dañinos para ciertas especies de pesca deportiva, como la trucha y el salmón.

b) Calcule la distancia aguas abajo a la cual el oxígeno se encuentra al mínimo. ¿Cuál es la concentración en dicha ubicación?

8.16 La concentración de bacterias contaminantes c en un lago disminuye de acuerdo con la ecuación

$$c = 70e^{-1.5t} + 25e^{-0.075t}$$

Determine el tiempo que se requiere para que la concentración de bacterias se reduzca a 9 con el uso de *a*) el método gráfico, y *b*) el método de Newton-Raphson, con un valor inicial de $t = 10$ y criterio de detención de 0.5%. Compruebe los resultados que obtenga.

8.17 Un cable en forma catenaria es aquel que cuelga entre dos puntos que no se encuentran sobre la misma línea vertical. Como se ilustra en la figura P8.17*a*, no está sujeto a más carga que su propio peso. Así, su peso (N/m) actúa como una carga uniforme por unidad de longitud a lo largo del cable. En la figura P8.17*b*, se ilustra un diagrama de cuerpo libre de una sección *AB*, donde T_A y T_B son las fuerzas de tensión en el extremo. Con base en los balances de fuerzas horizontal y vertical, se obtiene para el cable el siguiente modelo de ecuación diferencial:

$$\frac{d^2y}{dx^2} = \frac{w}{T_A}\sqrt{1+\left(\frac{dy}{dx}\right)^2}$$

Puede emplearse el cálculo para resolver esta ecuación para la altura y del cable como función de la distancia x,

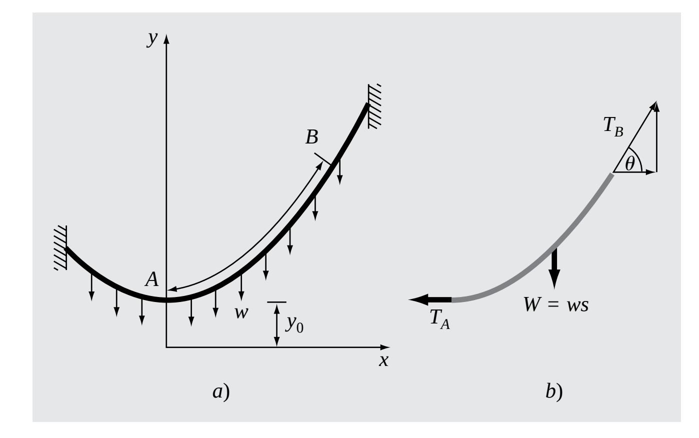


Figura P8.17
a) Fuerzas que actúan sobre una sección *AB* de un cable flexible que cuelga. La carga es uniforme a lo largo del cable (pero no uniforme por la distancia horizontal *x*). *b*) Diagrama de cuerpo libre de la sección *AB*.

$$y = \frac{T_A}{w}\cosh\left(\frac{w}{T_A}x\right) + y_0 - \frac{T_A}{w}$$

donde el coseno hiperbólico se calcula por medio de la ecuación:

$$\cosh x = \frac{1}{2}(e^x + e^{-x})$$

Utilice un método para calcular un valor para el parámetro T_A dados los valores de los parámetros $w = 10$ y $y_0 = 5$, de modo que el cable tenga una altura de $y = 15$ en $x = 50$.

8.18 En la figura P8.18*a* se muestra una viga uniforme sujeta a una carga distribuida uniformemente que crece en forma lineal. La ecuación para la curva elástica resultante es la siguiente (véase la figura P8.18*b*)

$$y = \frac{w_0}{120EIL}(-x^5 + 2L^2x^3 - L^4x) \qquad \text{(P8.18.1)}$$

Utilice el método de la bisección para determinar el punto de máxima deflexión (es decir, el valor de x donde $dy/dx = 0$). Después, sustituya este valor en la ecuación (P8.18.1) a fin de determinar el valor de la deflexión máxima. En sus cálculos, utilice los valores siguientes para los parámetros: $L = 450$ cm, $E = 50\,000$ kN/cm^2, $I = 30\,000$ cm^4 y $w_0 = 1.75$ kN/cm.

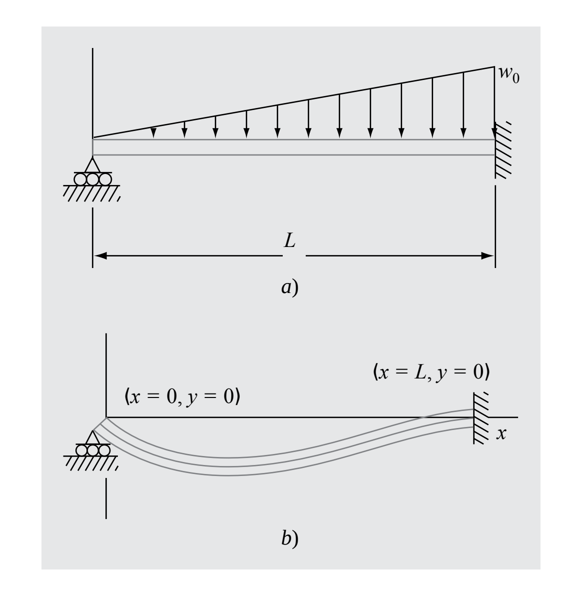


Figura P8.18

8.19 El desplazamiento de una estructura está definido por la ecuación siguiente para una oscilación amortiguada:

$$y = 8e^{-kt}\cos \omega t$$

donde $k = 0.5$ y $\omega = 3$.

a) Utilice el método gráfico para realizar una estimación inicial del tiempo que se requiere para que el desplazamiento disminuya a 4.

b) Emplee el método de Newton-Raphson para determinar la raíz con $\varepsilon_s = 0.01\%$.

c) Use el método de la secante para determinar la raíz con $\varepsilon_s = 0.01\%$.

8.20 La ecuación de Manning se puede escribir, para un canal rectangular abierto, como

$$Q = \frac{\sqrt{S}(BH)^{5/3}}{n(B+2H)^{2/3}}$$

donde Q = caudal [m^3/s], S = pendiente [m/m], H = profundidad [m] y n = coeficiente de rugosidad de Manning. Desarrolle un esquema de iteración de punto fijo para despejar H de esta ecuación, dados Q = 5, S = 0.0002, B = 20 y n = 0.03. Pruebe que su esquema converge para todos los valores iniciales mayores de cero o iguales a cero.

8.21 En ingeniería oceanográfica, la ecuación de una ola estacionaria reflejada en un puerto está dada por λ = 16, t = 12, v = 48:

$$h = h_0\left[\operatorname{sen}\left(\frac{2\pi x}{\lambda}\right)\cos\left(\frac{2\pi t v}{\lambda}\right) + e^{-x}\right]$$

Resuelva para el valor positivo más bajo de x, si $h = 0.4\ h_0$.

8.22 Suponga el lector que compra una pieza de equipo en \$20 000 como pago inicial y \$4 000 por año durante 6 años. ¿Qué tasa de interés estaría pagando? La fórmula que relaciona el valor presente P, los pagos anuales A, el número de años n y la tasa de interés i, es la que sigue:

$$A = P\frac{i(1+i)^n}{(1+i)^n - 1}$$

8.23 Muchos campos de la ingeniería requieren cálculos exactos de la población. Por ejemplo, los ingenieros de transporte quizás encuentren necesario determinar por separado la tendencia del crecimiento de una ciudad y la de los suburbios. La población del área urbana declina con el tiempo de acuerdo con la ecuación:

$$P_u(t) = P_{u,\text{máx}}e^{-k_u t} + P_{u,\text{mín}}$$

en tanto que la población suburbana crece según:

$$P_s(t) = \frac{P_{s,\text{ máx}}}{1 + [P_{s,\text{ máx}}/P_0 - 1]e^{-k_s t}}$$

donde $P_{u,\text{máx}}$, k_u, $P_{s,\text{máx}}$, P_0 y k_s son parámetros que se obtienen en forma empírica. Determine el tiempo y los valores correspondientes de $P_u(t)$ y $P_s(t)$ cuando los suburbios son 20% más grandes que la ciudad. Los valores de los parámetros son: $P_{u,\text{máx}}$ = 75 000, K_u = 0.045/año, $P_{u,\text{mín}}$ = 100 000 personas, $P_{s,\text{máx}}$ = 300 000 personas, P_0 = 10 000 personas, k_s = 0.08/año. Para obtener las soluciones utilice los métodos *a*) gráfico, *b*) de la falsa posición y *c*) de la secante modificada.

8.24 En la figura P8.24 se muestra una viga apoyada en forma sencilla que está cargada como se ilustra. Con el empleo de funciones de singularidad, el esfuerzo cortante a lo largo de la viga se expresa con la ecuación:

$$V(x) = 20[\langle x-0\rangle^1 - \langle x-5\rangle^1] - 15\langle x-8\rangle^0 - 57$$

Por definición, la función de singularidad se expresa del modo que sigue:

$$\langle x-a\rangle^n = \begin{cases}(x-a)^n & \text{cuando } x > a\\ 0 & \text{cuando } x \le a\end{cases}$$

Utilice un método numérico para encontrar el (los) punto(s) en los que el esfuerzo cortante sea igual a cero.

8.25 Con el uso de la viga apoyada en forma simple del problema 8.24, el momento a lo largo de ella, $M(x)$, está dado por:

$$M(x) = -10[\langle x-0\rangle^2 - \langle x-5\rangle^2] + 15\langle x-8\rangle^1 + 150\langle x-7\rangle^0 + 57x$$

Emplee un método numérico para encontrar el (los) punto(s) en los que el momento es igual a cero.

8.26 Con el uso de la viga con apoyo simple del problema 8.24, la pendiente a lo largo de ella está dada por:

$$\frac{du_y}{dx}(x) = \frac{-10}{3}[\langle x-0\rangle^3 - \langle x-5\rangle^3] + \frac{15}{2}\langle x-8\rangle^2 + 150\langle x-7\rangle^1 + \frac{57}{2}x^2 - 238.25$$

Utilice un método numérico para encontrar el (los) punto(s) donde la pendiente es igual a cero.

8.27 Para la viga con apoyo simple del problema 8.24, el desplazamiento a lo largo de ella está dado por la ecuación:

$$u_y(x) = \frac{-5}{6}[\langle x-0\rangle^4 - \langle x-5\rangle^4] + \frac{15}{6}\langle x-8\rangle^3 + 75\langle x-7\rangle^2 + \frac{57}{6}x^3 - 238.25x$$

a) Calcule el (los) punto(s) donde el desplazamiento es igual a cero.
b) ¿Cómo se usaría una técnica de localización de raíces para determinar la ubicación del desplazamiento mínimo?

8.28 Aunque no lo mencionamos en la sección 8.2, la ecuación (8.10) realmente es una expresión de *electroneutralidad*; es decir, de que las cargas positivas y negativas se deben equilibrar. Esto se puede ver más claramente expresándola como

$$[H^+] = [HCO_3^-] + 2[CO_3^{2-}] + [OH^-]$$

En otras palabras, las cargas positivas deben igualar a las cargas negativas. Entonces, cuando calcula el pH de un cuerpo de agua natural tal como un lago, usted debe tener en cuenta otros iones que pueden estar presentes. Para el caso en el que estos iones tengan su origen en sales no reactivas, las cargas netas negativas menos positivas debidas a estos iones se agrupan en una cantidad que se llama *alcalinidad*, y la ecuación se reformula como

$$Alk + [H^+] = [HCO_3^-] + 2[CO_3^{2-}] + [OH^-] \qquad \text{(P8.28.1)}$$

Donde Alk = alcalinidad (eq/L). Por ejemplo, la alcalinidad del Lago Superior es aproximadamente de 0.4×10^{-3} eq/L. Realice los mismos cálculos que en la sección 8.2 para calcular el pH del Lago Superior en 2008. Suponga que, igual que las gotas de lluvia, el lago está en equilibrio con el CO_2 atmosférico, pero tenga en cuenta la alcalinidad como en la ecuación (P8.28.1).

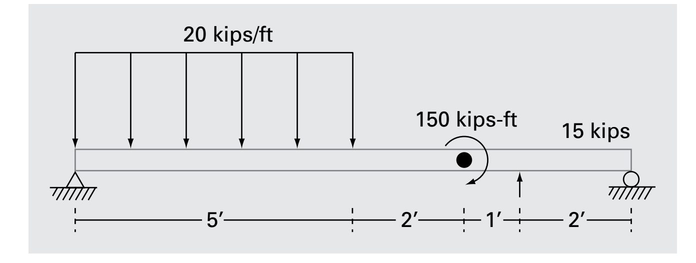


Figura P8.24

Ingeniería eléctrica

8.29 Ejecute el mismo cálculo que en la sección 8.3, pero determine el valor de L que se requiere para que el circuito disipe 1% de su valor original en $t = 0.05$ s, dado $R = 280\ \Omega$ y $C = 10^{-4}$ F. Emplee *a*) un enfoque gráfico, *b*) la bisección y *c*) software para encontrar raíces, como Solver de Excel o la función `fzero` de MATLAB, o la raíz de la función de Mathcad.

8.30 La ecuación $i = 9e^{-t}$ sen $(2\pi t)$, describe una corriente oscilatoria en un circuito eléctrico, donde t se expresa en segundos. Determine todos los valores de t de modo que $i = 3.5$.

8.31 La resistividad ρ de un lubricante de sílice se basa en la carga q en un electrón, la densidad del electrón n y la movilidad del electrón μ. La densidad del electrón está dada en términos de la densidad del lubricante N y la densidad intrínseca de acarreo n_i. La movilidad del electrón está descrita por la temperatura T, la temperatura de referencia T_0 y la movilidad de referencia μ_0. Las ecuaciones que se requieren para calcular la resistividad son las siguientes:

$$\rho = \frac{1}{qn\mu}$$

donde

$$n = \frac{1}{2}\left(N + \sqrt{N^2 + 4n_i^2}\right) \qquad \text{y} \qquad \mu = \mu_0\left(\frac{T}{T_0}\right)^{-2.42}$$

Determine N, dado que $T_0 = 300$ K, $T = 1000$ K, $\mu_0 = 1300$ cm^2 (V s)$^{-1}$, $q = 1.6 \times 10^{-19}$ C, $n_i = 6.21 \times 10^9$ cm^{-3} y un valor deseable de $\rho = 6 \times 10^6$ V s cm/C. Use los métodos *a*) bisección y *b*) la secante modificada.

8.32 Una carga total Q se encuentra distribuida en forma uniforme alrededor de un conductor en forma de anillo con radio a. Una carga q se localiza a una distancia x del centro del anillo (véase la figura P8.32). La fuerza que el anillo ejerce sobre la carga está dada por la ecuación

$$F = \frac{1}{4\pi e_0}\frac{qQx}{(x^2 + a^2)^{3/2}}$$

donde $e_0 = 8.85 \times 10^{-12}$ C^2/(N m^2). Encuentre la distancia x donde la fuerza es de 1 N, si q y Q son 2×10^{-5} C para un anillo con un radio de 0.9 m.

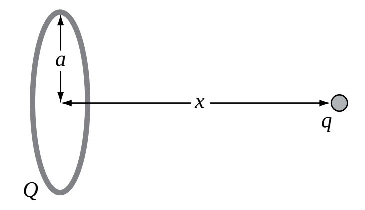


Figura P8.32

8.33 En la figura P8.33 se muestra un circuito con una resistencia, un inductor y un capacitor en paralelo. Para expresar la impedancia del sistema se emplean las leyes de Kirchhoff, así:

$$\frac{1}{Z} = \sqrt{\frac{1}{R^2} + \left(\omega C - \frac{1}{\omega L}\right)^2}$$

donde Z = impedancia (Ω) y ω = frecuencia angular. Encuentre la ω que da como resultado una impedancia de 75 Ω, con el uso tanto del método de la bisección como el de la falsa posición, con valores iniciales de 1 y 1000 y los parámetros siguientes: $R = 225\ \Omega$, $C = 0.6 \times 10^{-6}$ F y $L = 0.5$ H. Determine cuántas iteraciones son necesarias con cada técnica a fin de encontrar la respuesta con $\varepsilon_s = 0.1\%$. Utilice el enfoque gráfico para explicar cualesquiera dificultades que surjan.

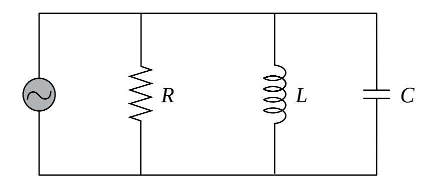


Figura P8.33

Ingeniería mecánica y aeroespacial

8.34 Más allá de la *ecuación de Colebrook*, hay disponibles otras relaciones como el *factor de fricción de Fanning f* para estimar la fricción en tubos. El factor de fricción de Fanning depende de cierto número de parámetros relacionados con el tamaño del tubo y el fluido, que pueden representarse con otra cantidad adimensional, el *número de Reynolds* Re. Una fórmula que pronostica el valor de f dado Re es la *ecuación de Von Karman*.

$$\frac{1}{\sqrt{f}} = 4\log_{10}\left(\text{Re}\sqrt{f}\right) - 0.4$$

Valores comunes del número de Reynolds para flujo turbulento son 10000 a 500000, y del factor de fricción de Fanning son 0.001 a 0.01. Desarrolle una función que utilice el método de bisección con objeto de resolver cuál sería el factor de fricción de Fanning f, dado un valor de Re proporcionado por el usuario que esté entre 2500 y 1000000. Diseñe la función de modo que se garantice que el error absoluto en el resultado sea de $E_{a,d} < 0.000005$.

8.35 Los sistemas mecánicos reales involucran la deflexión de resortes no lineales. En la figura P8.35 se ilustra una masa m que se libera por una distancia h sobre un resorte no lineal. La fuerza de resistencia F del resorte está dada por la ecuación

$$F = -\left(k_1 d + k_2 d^{3/2}\right)$$

Es posible usar la conservación de la energía para demostrar que

$$0 = \frac{2k_2 d^{5/2}}{5} + \frac{1}{2}k_1 d^2 - mgd - mgh$$

Despeje d, dados los siguientes valores de parámetros: $k_1 = 40\,000$ g/s^2, $k_2 = 40$ g/(s^2 m$^{0.5}$), $m = 95$ g, $g = 9.81$ m/s^2 y $h = 0.43$ m.

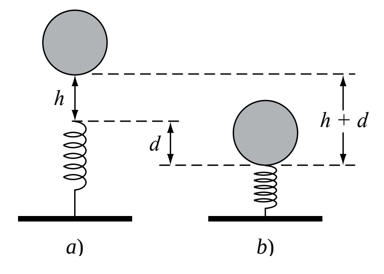


Figura P8.35

8.36 Los ingenieros mecánicos, así como los de otras especialidades, utilizan mucho la termodinámica para realizar su trabajo. El siguiente polinomio se emplea para relacionar el calor específico a presión cero del aire seco, c_p kJ/(kg K), a temperatura (K):

$$c_p = 0.99403 + 1.671 \times 10^{-4} T + 9.7215 \times 10^{-8} T^2 - 9.5838 \times 10^{-11} T^3 + 1.9520 \times 10^{-14} T^4$$

Determine la temperatura que corresponda a un calor específico de 1.2 kJ/(kg K).

8.37 En ciertas ocasiones, los ingenieros aeroespaciales deben calcular las trayectorias de proyectiles, como cohetes. Un problema parecido tiene que ver con la trayectoria de una pelota que se lanza. Dicha trayectoria está definida por las coordenadas (x, y), como se ilustra en la figura P8.37. La trayectoria se modela con la ecuación

$$y = (\tan\theta_0)x - \frac{g}{2v_0^2\cos^2\theta_0}x^2 + y_0$$

Calcule el ángulo inicial θ_0, apropiado si la velocidad inicial $v_0 = 20$ m/s y la distancia x al *catcher* es de 40 m. Obsérvese que la pelota sale de la mano del lanzador con una elevación $y_0 = 1.8$ m, y el *catcher* la recibe a 1 m. Exprese el resultado final en grados. Para g, utilice un valor de 9.81 m/s^2, y emplee el método gráfico para elegir valores iniciales.

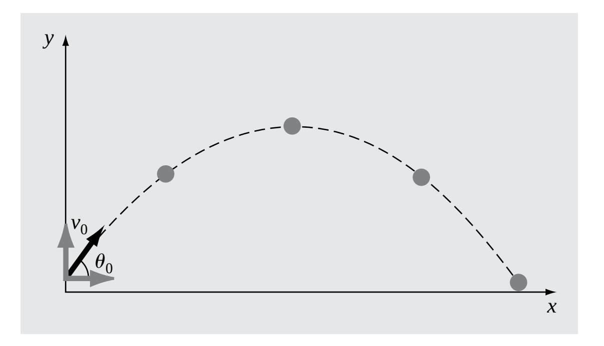


Figura P8.37

8.38 La forma general para un campo tensorial de tres dimensiones es la siguiente:

$$\begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_{zz} \end{bmatrix}$$

en la que los términos en la diagonal principal representan esfuerzos a la tensión o a la compresión, y los términos fuera de la diagonal representan los esfuerzos cortantes. Un campo tensorial (en MPa) está dado por la matriz que sigue:

$$\begin{bmatrix} 10 & 14 & 25 \\ 14 & 7 & 15 \\ 25 & 15 & 16 \end{bmatrix}$$

Para resolver cuáles son los esfuerzos principales, es necesario construir la matriz siguiente (de nuevo en MPa):

$$\begin{bmatrix} 10-\sigma & 14 & 25 \\ 14 & 7-\sigma & 15 \\ 25 & 15 & 16-\sigma \end{bmatrix}$$

σ_1, σ_2 y σ_3 se obtienen con la ecuación

$$\sigma^3 - I\sigma^2 + II\sigma - III = 0$$

donde

$$I = \sigma_{xx} + \sigma_{yy} + \sigma_{zz}$$

$$II = \sigma_{xx}\sigma_{yy} + \sigma_{xx}\sigma_{zz} + \sigma_{yy}\sigma_{zz} - \sigma_{xy}^2 - \sigma_{xz}^2 - \sigma_{yz}^2$$

$$III = \sigma_{xx}\sigma_{yy}\sigma_{zz} - \sigma_{xx}\sigma_{yz}^2 - \sigma_{yy}\sigma_{xz}^2 - \sigma_{zz}\sigma_{xy}^2 + 2\sigma_{xy}\sigma_{xz}\sigma_{yz}$$

I, *II* y *III* se conocen como las invariantes de esfuerzos. Encuentre σ_1, σ_2 y σ_3 por medio de una técnica de localización de raíces.

8.39 La velocidad vertical de un cohete se calcula con la fórmula que sigue:

$$v = u\ln\frac{m_0}{m_0 - qt} - gt$$

donde v = velocidad vertical, u = velocidad con la que se expele el combustible, en relación con el cohete, m_0 = masa inicial del cohete en el momento $t = 0$, q = tasa de consumo de combustible y g = aceleración de la gravedad hacia abajo (se supone constante e igual a 9.81 m/s^2). Si u = 2 200 m/s, m_0 = 160 000 kg y q = 2 680 kg/s, calcule el momento en que v = 1 000 m/s. (*Sugerencia:* El valor de t se encuentra entre 10 y 50 s.) Calcule el resultado de modo que esté dentro de 1% del valor verdadero. Compruebe su respuesta.

8.40 El ángulo de fase ϕ entre la vibración forzada que ocasiona el camino rugoso y el movimiento del carro, está dada por la ecuación:

$$\tan\phi = \frac{2(c/c_c)(\omega/p)}{1-(\omega/p)^2}$$

Como ingeniero mecánico, le gustaría saber si existen casos en que $\phi = \omega/2 - 1$. Utilice los otros parámetros de la sección con objeto de plantear la ecuación como un problema de cálculo de raíces, y resuélvala para ω.

8.41 Se mezclan dos fluidos con temperatura diferente de modo que alcanzan la misma temperatura. La capacidad calorífica del fluido A está dada por:

$$c_p = 3.381 + 1.804 \times 10^{-2} T - 4.300 \times 10^{-6} T^2$$

y la capacidad calorífica del fluido B se obtiene con:

$$c_p = 8.592 + 1.290 \times 10^{-1} T - 4.078 \times 10^{-5} T^2$$

donde c_p se expresa en unidades de cal/mol K, y T está en unidades de K. Obsérvese que

$$\Delta H = \int_{T_1}^{T_2} c_p\,dT$$

El fluido A entra al mezclador a 400°C, y el B a 600°C. Al entrar al mezclador hay lo doble de fluido A que B. ¿A qué temperatura salen los dos fluidos del mezclador?

8.42 Un compresor opera a una razón de compresión R_c de 3.0 (esto significa que la presión del gas en la salida es tres veces mayor que en la entrada). Los requerimientos de energía del compresor H_p se determinan por medio de la ecuación que se da a continuación. Suponga que los requerimientos de energía del compresor son exactamente iguales a zRT_1/MW, y encuentre la eficiencia politrópica n del compresor. El parámetro z es la compresibilidad del gas en las condiciones de operación del compresor, R es la constante de los gases, T_1 es la temperatura del gas en la entrada del compresor y MW es el peso molecular del gas.

$$\text{HP} = \frac{zRT_1}{\text{MW}}\frac{n}{n-1}\left(R_c^{(n-1)/n} - 1\right)$$

8.43 En los envases térmicos que se ilustran en la figura P8.43, el compartimiento interior está separado del medio por medio de vacío. Hay una cubierta exterior alrededor de los envases. Esta cubierta está separada de la capa media por una capa delgada de aire. La superficie de afuera de la cubierta exterior está en contacto con el aire del ambiente. La transferencia de calor del compartimiento interior a la capa siguiente q_1 sólo ocurre por radiación (ya que el espacio se encuentra vacío). La transferencia de calor entre la capa media y la cubierta exterior q_2 es por convección en un espacio pequeño. La transferencia de calor de la cubierta exterior hacia el aire q_3 sucede por convección natural. El flujo de calor desde cada región de los envases debe ser igual, es decir, $q_1 = q_2 = q_3$. Encuentre las temperaturas T_1 y T_2 en estado estable. T_0 es de 500°C y T_3 = 25°C.

$$q_1 = 10^{-9}[(T_0 + 273)^4 - (T_1 + 273)^4]$$
$$q_2 = 4(T_1 - T_2)$$
$$q_3 = 1.3(T_2 - T_3)^{4/3}$$

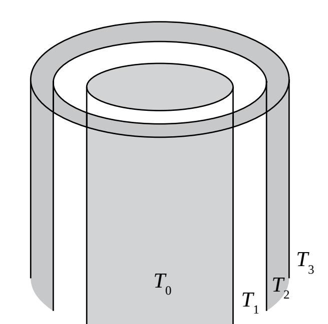


Figura P8.43

8.44 La figura P8.44 ilustra tres almacenamientos conectados por medio de tubos circulares. Los tubos están hechos de hierro fundido recubierto con asfalto (ε = 0.0012 m), y tienen las características siguientes:

Tubo	1	2	3
Longitud, m	1 800	500	1 400
Diámetro, m	0.4	0.25	0.2
Flujo, m^3/s	?	0.1	?

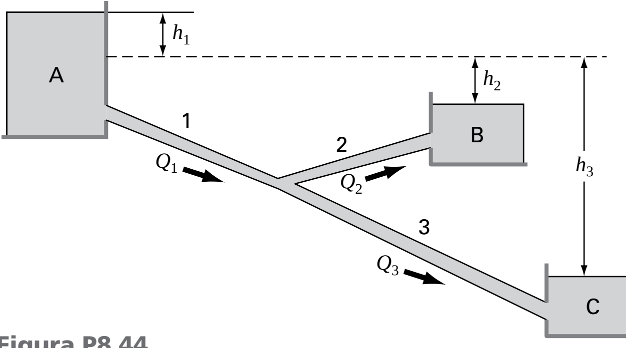


Figura P8.44

Si las elevaciones de la superficie del agua en los almacenamientos A y C son de 200 m y 172.5 m, respectivamente, determine la elevación que alcanza en el almacenamiento B y los flujos en los tubos 1 y 3. Obsérvese que la viscosidad cinemática del agua es de 1×10^{-6} m^2/s, y utilice la *ecuación de Colebrook* para obtener el factor de fricción (consulte el problema 8.13).

8.45 Un fluido se bombea en la red de tubos que se muestra en la figura P8.45. En estado estacionario, se cumplen los balances de flujo siguientes:

$$Q_1 = Q_2 + Q_3$$
$$Q_3 = Q_4 + Q_5$$
$$Q_5 = Q_6 + Q_7$$

donde Q_i = flujo en el tubo i [m^3/s]. Además, la caída de presión alrededor de los tres lazos en los que el flujo es hacia la derecha debe ser igual a cero. La caída de presión en cada tramo de tubo circular se calcula por medio de la ecuación:

$$\Delta P = \frac{16}{\pi^2}\frac{fL\rho}{2D^5}Q^2$$

donde ΔP = caída de presión (Pa), f = factor de fricción (adimensional), L = longitud del tubo (m), ρ = densidad del fluido (kg/m^3) y D = diámetro del tubo (m). Escriba un programa (o desarrolle un algoritmo en algún paquete de software de matemáticas) que permita calcular el flujo en cada tramo de tubo, dado que $Q_1 = 1$ m^3/s y ρ = 1.23 kg/m^3. Todos los tubos tienen D = 500 mm y f = 0.005. Las longitudes de los tubos son: $L_3 = L_5 = L_8 = L_9$ = 2 m; $L_2 = L_4 = L_6$ = 4 m, y L_7 = 8 m.

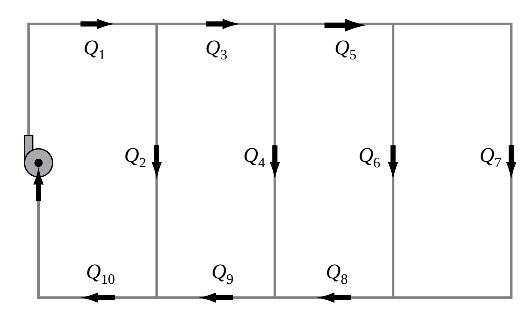


Figura P8.45

8.46 Repita el problema 8.45, pero incorpore el hecho de que el factor de fricción se calcula con la *ecuación de Von Karman*, que es:

$$\frac{1}{\sqrt{f}} = 4\log_{10}(\text{Re}\sqrt{f}) - 0.4$$

donde Re = número de Reynolds

$$\text{Re} = \frac{\rho VD}{\mu}$$

donde V = velocidad del fluido en el tubo (m/s) y μ = viscosidad dinámica (N·s/m^2). Obsérvese que para un tubo circular, $V = 4Q/\pi D^2$. Asimismo, suponga que el fluido tiene una viscosidad de 1.79×10^{-5} N·s/m^2.

8.47 Sobre el transbordador espacial, al despegar de la plataforma, actúan cuatro fuerzas, las que se muestran en el diagrama de cuerpo libre (véase la figura P8.47). El peso combinado de los dos cohetes

de combustible sólido y del tanque exterior de éste, es de $W_B = 1.663 \times 10^6$ lb. El peso del orbitador con carga completa es de $W_S = 0.23 \times 10^6$ lb. El empuje combinado de los dos cohetes de combustible sólido es $T_B = 5.30 \times 10^6$ lb. El empuje combinado de los tres motores de combustible líquido del orbitador es de $T_S = 1.125 \times 10^6$ lb.

Al despegar, el empuje del motor del orbitador se dirige con un ángulo θ para hacer que el momento resultante que actúa sobre el conjunto de la nave (tanque exterior, cohetes de combustible sólido y orbitador) sea igual a cero. Con el momento resultante igual a cero, la nave no giraría sobre su centro de gravedad *G* al despegar. Con estas fuerzas, la nave experimentará una fuerza resultante con componentes en dirección vertical y horizontal. La componente vertical de la fuerza resultante, es la que permite que la nave despegue de la plataforma y vuele verticalmente. La componente horizontal de la fuerza resultante hace que la nave vuele en forma horizontal. El momento resultante que actúa sobre la nave será igual a cero cuando θ se ajusta al valor apropiado. Si este ángulo no se ajusta en forma adecuada y hubiera algún momento que actuara sobre la nave, ésta tendería a girar alrededor de su centro de gravedad.

a) Resuelva el empuje del orbitador T_S en las componentes horizontal y vertical, y después sume los momentos respecto del punto *G*, centro de gravedad de la nave. Iguale a cero la ecuación del momento resultante. Ahora, ésta puede resolverse para el valor de θ que se requiere durante el despegue.

b) Obtenga una ecuación para el momento resultante que actúa sobre la nave en términos del ángulo θ. Grafique el momento resultante como función del ángulo θ en el rango de –5 radianes a +5 radianes.

c) Escriba un programa de computadora para resolver para el ángulo θ por medio del método de Newton para encontrar la raíz de la ecuación del momento resultante. Con el empleo de la gráfica, elija un valor inicial para la raíz de interés. Interrumpa las iteraciones cuando el valor de θ ya no mejore con cinco cifras significativas.

d) Repita el programa para el peso de la carga mínima del orbitador, que es $W_S = 195\,000$ lb.

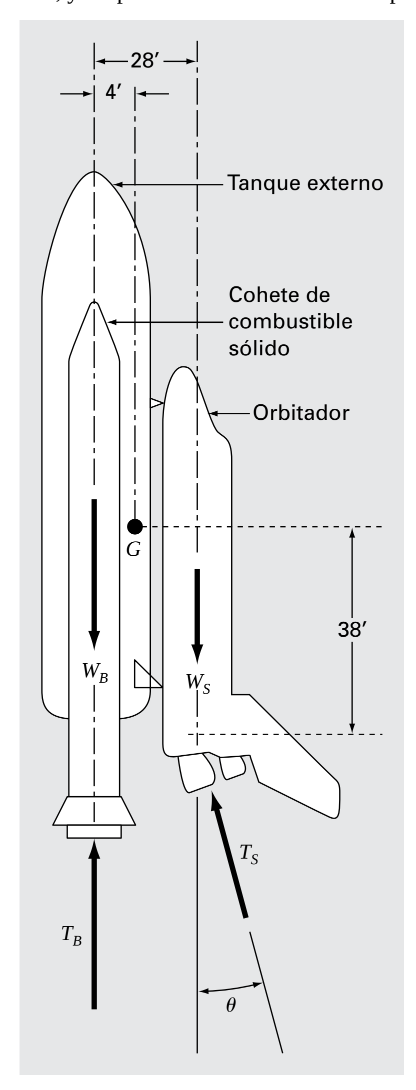


Figura P8.47

8.48 Determinar la velocidad de partículas que se sedimentan a través de un fluido es de gran importancia en muchas áreas de la ingeniería y de la ciencia. Tales cálculos dependen del régimen de flujo representado por el *número de Reynolds*, adimensional,

$$\text{Re} = \frac{\rho d v}{\mu} \tag{P8.48.1}$$

donde ρ = densidad del fluido (kg/m^3), d = diámetro de la partícula (m), v = velocidad de sedimentación de la partícula (m/s) y μ = viscosidad dinámica del fluido (Ns/m^2). En condiciones de flujo laminar (Re < 0.1), la velocidad de sedimentación de una partícula esférica se puede calcular con la siguiente fórmula basada en la *ley de Stokes*:

$$v = \frac{g}{18}\left(\frac{\rho_s - \rho}{\mu}\right)d^2 \tag{P8.48.2}$$

donde g = constante gravitacional (= 9.81 m/s^2) y ρ_s = densidad de la partícula (kg/m^3). Para condiciones de flujo turbulento (es decir, números de Reynolds mayores), se puede usar un procedimiento alternativo basado en la siguiente fórmula:

$$v = \sqrt{\frac{4g(\rho_s - \rho)d}{3C_D\rho}} \tag{P8.48.3}$$

donde C_D = coeficiente de arrastre, que depende del número de Reynolds, como en

$$C_D = \frac{24}{\text{Re}} + \frac{3}{\sqrt{\text{Re}}} + 0.34 \tag{P8.48.4}$$

a) Combine las ecuaciones (P8.48.2), (P8.48.3) y (P8.48.4) para expresar la determinación de *v* como una raíz de las ecuaciones del problema. Es decir, exprese la fórmula combinada en el formato $f(v) = 0$.

b) Use el método de secante modificada con $\delta = 10^{-3}$ y $\varepsilon_s = 0.05\%$ para determinar *v* para una partícula esférica de hierro sedimentándose en agua, donde $d = 200\ \mu$m, $\rho = 1$ g/cm^3, $\rho_s = 7.874$ g/cm^3 y $\mu = 0.014$ g/(cm·s). Utilice la ecuación (P8.48.2) para generar su valor inicial.

c) Con base en el resultado de *b*), calcule el número de Reynolds y el coeficiente de arrastre, y use este último para confirmar que el régimen de flujo no es laminar.

d) Desarrolle una solución con iteración de punto fijo para las condiciones descritas en *b*).

e) Use un procedimiento gráfico para ilustrar que la formulación desarrollada en *d*) convergirá para cualquier valor positivo.

EPÍLOGO: PARTE DOS

PT2.4 ALTERNATIVAS

La tabla PT2.3 proporciona un resumen de las alternativas para la solución de las raíces de ecuaciones algebraicas y trascendentes. Aunque los métodos gráficos consumen tiempo, ofrecen cierto conocimiento sobre el comportamiento de la función y son útiles para identificar valores iniciales y problemas potenciales como el de las raíces múltiples. Por lo tanto, si el tiempo lo permite, un bosquejo rápido (o mejor aún, una gráfica computarizada) brindará información valiosa sobre el comportamiento de la función.

Los métodos numéricos se dividen en dos grandes categorías: métodos cerrados y abiertos. Los primeros requieren dos valores iniciales que estén a ambos lados de la raíz, para acotarla. Este "acotamiento" se mantiene en tanto se aproxima a la solución; así, dichas técnicas son siempre convergentes. Sin embargo, se debe pagar un precio por esta propiedad, la velocidad de convergencia es relativamente lenta.

Las técnicas abiertas difieren de los métodos cerrados inicialmente en que usan la información de un solo punto (o dos valores que no necesitan acotar a la raíz para extrapolar a una nueva aproximación de la misma). Esta propiedad es una espada de dos filos. Aunque llevan a una rápida convergencia, existe la posibilidad de que la solución diverja. En general, la convergencia con técnicas abiertas es parcialmente dependiente de la calidad del valor inicial y de la naturaleza de la función. Cuanto más cerca esté el valor inicial de la raíz verdadera, los métodos convergerán más rápido.

De las técnicas abiertas, el método estándar de Newton-Raphson se utiliza con frecuencia por su propiedad de convergencia cuadrática. Sin embargo, su mayor deficiencia es que requiere que la derivada de la función se obtenga en forma analítica. Con algunas funciones se vuelve impráctico.

Tabla PT2.3 Comparación de las características de los métodos alternativos para encontrar raíces de ecuaciones algebraicas y trascendentes. Las comparaciones se basan en la experiencia general y no toman en cuenta el comportamiento de funciones específicas.

Método	Tipo	Valores	Convergencia	Estabilidad	Programación	Comentarios
Directo	Analítico	—	—	—		
Gráfico	Visual	—	—	—	—	Impreciso
Bisección	Cerrado	2	Lenta	Siempre	Fácil	
Falsa posición	Cerrado	2	Lenta/media	Siempre	Fácil	
PF modificada	Cerrado	2	Media	Siempre	Fácil	
Iteración punto fijo	Abierto	1	Lenta	Posiblemente divergente	Fácil	
Newton-Raphson	Abierto	1	Rápida	Posiblemente divergente	Fácil	Necesita evaluación de $f'(x)$
Newton-Raphson modificado	Abierto	1	Rápida (múltiple), media (única)	Posiblemente divergente	Fácil	Necesita evaluación de $f'(x)$ y $f''(x)$
Secante	Abierto	2	Mediana/rápida	Posiblemente divergente	Fácil	Condiciones iniciales no tienen que abarcar la raíz
Secante modificado	Abierto	1	Media/rápida	Posiblemente divergente	Fácil	Robusto
Brent	Híbrido	1 o 2	Media	Siempre (para 2 condiciones)	Moderada	
Müller	Polinomios	2	Media/rápida	Posiblemente divergente	Moderada	
Bairstow	Polinomios	2	Rápida	Posiblemente divergente	Moderada	

En dichos casos, el método de la secante, que emplea una representación en diferencias finitas de la derivada, proporciona una alternativa viable. Debido a la aproximación en diferencias finitas, la velocidad de convergencia del método de la secante es al principio más lento que el método de Newton-Raphson. Sin embargo, conforme se refina la estimación de la raíz, la aproximación por diferencias se vuelve una mejor representación de la derivada verdadera y, en consecuencia, se acelera rápidamente la convergencia. Se puede usar la técnica modificada de Newton-Raphson y así obtener una rápida convergencia para raíces múltiples. Sin embargo, dicha técnica requiere una expresión analítica tanto para la primera como para la segunda derivada.

Son especialmente interesantes los métodos híbridos que combinan la confiabilidad del acotamiento con la rapidez de los métodos abiertos. El método de Brent hace esto combinando la bisección con diversos métodos abiertos. Todos los métodos numéricos son fáciles de programar en computadoras y requieren de un tiempo mínimo para determinar una sola raíz. Sobre esta base, usted podría concluir que los métodos simples como el de bisección resultarían suficientemente buenos para fines prácticos. Lo anterior será cierto si usted se interesa exclusivamente en determinar sólo una vez la raíz de una ecuación. Pero hay muchos casos en ingeniería donde se requiere la localización de muchas raíces y donde la rapidez se vuelve importante. En tales casos, los métodos lentos consumen mucho tiempo y son por lo tanto costosos. Por otro lado, los métodos abiertos (más rápidos) pueden llegar a diverger, y los retardos que esto implica pueden también ser costosos. Algunos algoritmos de cómputo intentan conjugar las ventajas de ambas técnicas, al emplear inicialmente un método cerrado para aproximar la raíz, y después cambiar a un método abierto que mejore la estimación con rapidez. Ya sea que se utilice un solo procedimiento o una combinación, la búsqueda de convergencia y velocidad es fundamental para la elección de una técnica de localización de raíces.

PT2.5 RELACIONES Y FÓRMULAS IMPORTANTES

La tabla PT2.4 resume la información importante que se presentó en la parte dos. Dicha tabla se puede consultar para un acceso rápido de relaciones y fórmulas importantes.

PT2.6 MÉTODOS AVANZADOS Y REFERENCIAS ADICIONALES

En el presente texto los métodos se han concentrado en determinar una sola raíz real de una ecuación algebraica o trascendente, considerando un conocimiento previo de su localización aproximada. Además, se han descrito también métodos que se hallan expresamente diseñados para determinar las raíces reales y complejas de polinomios. Referencias adicionales sobre el tema son Ralston y Rabinowitz (1978) y Carnahan, Luther y Wilkes (1969).

Además de los métodos de Müller y de Bairstow, existen varias técnicas disponibles para determinar todas las raíces de polinomios. En particular, el *algoritmo de diferencia del cociente* (*QD*) (Henrici, 1964, y Gerald y Wheatley, 2004) determina todas las raíces sin tener valores iniciales. Ralston y Rabinowitz (1978) y Carnahan, Luther y Wilkes (1969) contienen un análisis de este método, así como de otras técnicas para la localización de raíces de polinomios. Como se analiza en el texto, los métodos de *Jenkins-Traub* y de *Laguerre* son de uso frecuente.

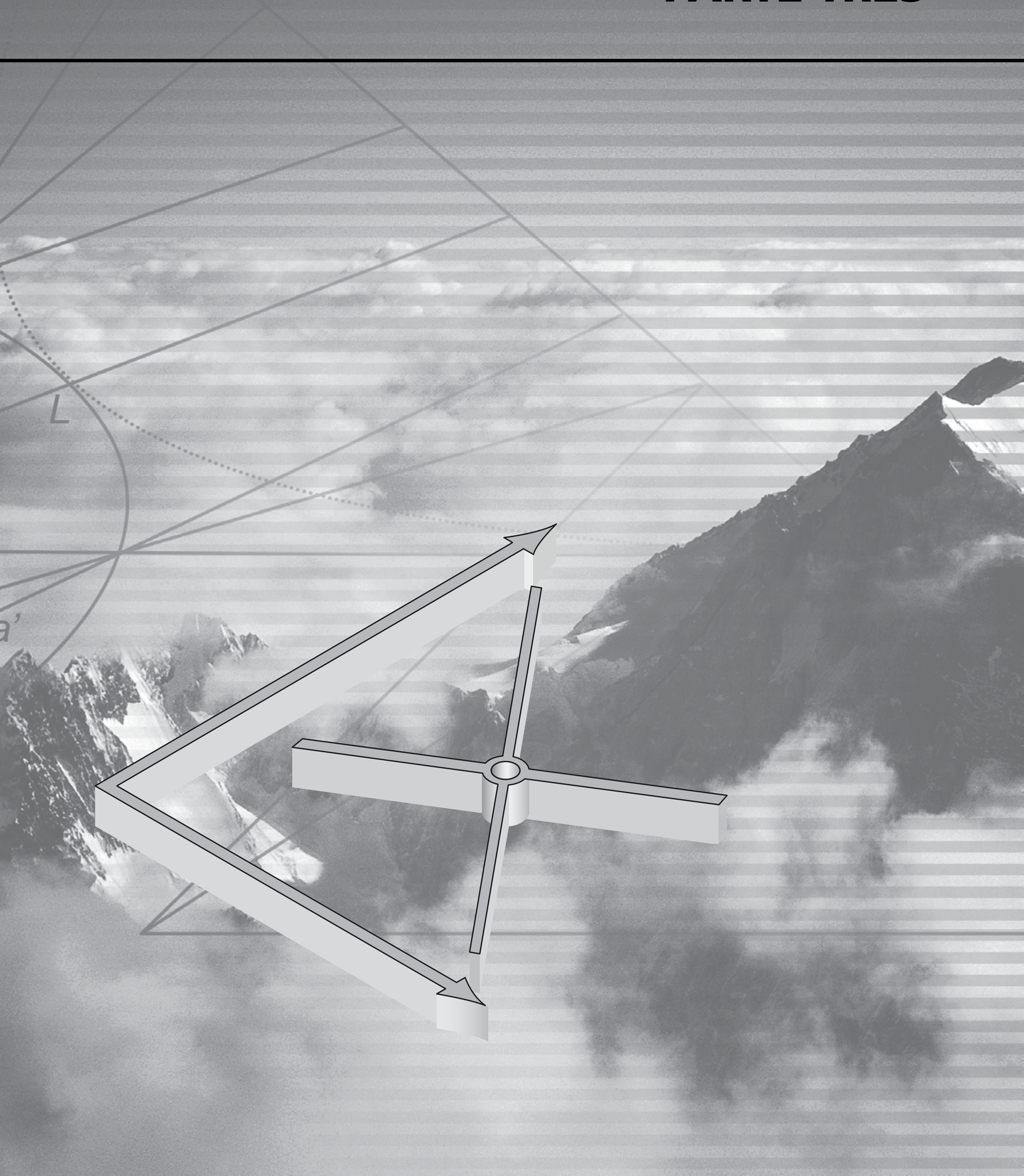
En resumen, lo anterior lleva la intención de proporcionarle nuevos caminos para una exploración más profunda del tema. Además, todas las referencias anteriores ofrecen descripciones de las técnicas básicas cubiertas en la parte dos. Le recomendamos que consulte esas fuentes alternativas con el objetivo de ampliar su comprensión de los métodos numéricos para la localización de raíces.[1]

[1] Aquí sólo se menciona el autor de los libros citados. Se puede encontrar una bibliografía completa al final de este texto.

Tabla PT2.4 Resumen de información importante presentada en la parte dos.

Método	Formulación	Interpretación gráfica	Errores y criterios de terminación
		Métodos cerrados:	
Bisección	$x_r = \dfrac{x_l + x_u}{2}$ Si $f(x_l)f(x_r) < 0,\ x_u = x_r$ $f(x_l)f(x_r) > 0,\ x_l = x_r$	$f(x)$; Raíz; x_l; x_u; x; L; $L/2$; $L/4$	Criterio de terminación: $\left\lvert \dfrac{x_r^{\text{nuevo}} - x_r^{\text{anterior}}}{x_r^{\text{nuevo}}} \right\rvert 100\% \le \epsilon_s$
Falsa posición	$x_r = x_u - \dfrac{f(x_u)(x_l - x_u)}{f(x_l) - f(x_u)}$ Si $f(x_l)f(x_r) < 0,\ x_u = x_r$ $f(x_l)f(x_r) > 0,\ x_l = x_r$	$f(x)$; Cuerda; x_r; x_l; x_u; x	Criterio de terminación: $\left\lvert \dfrac{x_r^{\text{nuevo}} - x_r^{\text{anterior}}}{x_r^{\text{nuevo}}} \right\rvert 100\% \le \epsilon_s$
Newton-Raphson	$x_{i+1} = x_i - \dfrac{f(x_i)}{f'(x_i)}$	$f(x)$; Tangente; x_{i+1}; x_i; x	Criterio de terminación: $\left\lvert \dfrac{x_{i+1} - x_i}{x_{i+1}} \right\rvert 100\% \le \epsilon_s$ Error: $E_{i+1} = 0(E_i^2)$
Secante	$x_{i+1} = x_i - \dfrac{f(x_i)(x_{i-1} - x_i)}{f(x_{i-1}) - f(x_i)}$	$f(x)$; x_{i+1}; x_i; x_{i-1}; x	Criterio de terminación: $\left\lvert \dfrac{x_{i+1} - x_i}{x_{i+1}} \right\rvert 100\% \le \epsilon_s$

PARTE TRES



ECUACIONES ALGEBRAICAS LINEALES

PT3.1 MOTIVACIÓN

En la parte dos, determinamos el valor de x que satisface una única ecuación, $f(x) = 0$. Ahora, nos ocuparemos de determinar los valores $x_1, x_2, \ldots, x_n$ que en forma simultánea satisfacen un sistema de ecuaciones

$$
\begin{aligned}
f_1(x_1, x_2, \ldots, x_n) &= 0 \\
f_2(x_1, x_2, \ldots, x_n) &= 0 \\
&\vdots \\
f_n(x_1, x_2, \ldots, x_n) &= 0
\end{aligned}
$$

Tales sistemas pueden ser lineales o no lineales. En la parte tres, trataremos con *ecuaciones algebraicas lineales*, que tienen la forma general

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
&\vdots \\
a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n
\end{aligned}
\qquad \text{(PT3.1)}
$$

donde las a son los coeficientes constantes, las b son constantes y n es el número de ecuaciones. Todas las demás ecuaciones son no lineales. Los sistemas no lineales se analizaron en el capítulo 6, aunque se volverán a estudiar brevemente en el capítulo 9.

PT3.1.1 Métodos sin computadora para resolver sistemas de ecuaciones

Si son pocas ecuaciones ($n \leq 3$), las ecuaciones lineales (y algunas veces las no lineales) pueden resolverse con rapidez mediante técnicas simples. Algunos de estos métodos se revisarán al inicio del capítulo 9. Sin embargo, con cuatro o más ecuaciones, la solución se vuelve laboriosa y debe usarse una computadora. Históricamente, la incapacidad para resolver a mano los sistemas de ecuaciones más grandes ha limitado el alcance de problemas por resolver en muchas aplicaciones de ingeniería.

Antes de las computadoras, las técnicas para resolver ecuaciones algebraicas lineales consumían mucho tiempo y eran poco prácticas. Esos procedimientos restringieron la creatividad debido a que con frecuencia los métodos eran difíciles de implementar y entender. Como resultado, las técnicas se sobreenfatizaron, a expensas de otros aspectos del proceso de solución de problemas como la formulación y la interpretación (recuerde la figura PT1.1 y el análisis respectivo).

El surgimiento de las computadoras fácilmente accesibles hizo posible y práctico resolver grandes sistemas de ecuaciones algebraicas lineales simultáneas. Así, puede enfrentar ejemplos y problemas más complicados. Además, cuenta con más tiempo para usar sus habilidades creativas, ya que pondrá mayor énfasis en la formulación del problema y en la interpretación de la solución.

PT3.1.2 Ecuaciones algebraicas lineales y la práctica en ingeniería

Muchas de las ecuaciones fundamentales en ingeniería se basan en las leyes de conservación (recuerde la tabla 1.1). Entre algunas cantidades conocidas que se someten a tales leyes están la masa, la energía y el *momentum*. En términos matemáticos, estos principios nos conducen a ecuaciones de balance o de continuidad que relacionan el *comportamiento* del sistema, al representarlo por los *niveles* o *respuesta* de la cantidad que se modela con las *propiedades* o *características* del sistema, y por los *estímulos* externos o *funciones forzadas* que actúan sobre el sistema.

Por ejemplo, el principio de conservación de la masa se utiliza para formular un modelo de una serie de reactores químicos (figura PT3.1*a*). En este caso, la cantidad que habrá de modelarse es la masa de las sustancias químicas en cada reactor. Las propiedades del sistema son la reacción característica de la sustancia química, los tamaños de los reactores y las velocidades de flujo. Las funciones forzadas son las velocidades de suministro de las sustancias químicas hacia el sistema.

En la parte dos, usted observó cómo sistemas de un solo componente dan por resultado una sola ecuación que puede resolverse mediante técnicas de localización de raíces. Los sistemas con multicomponentes resultan en un sistema de ecuaciones matemáticas que deben resolverse de manera simultánea. Las ecuaciones están relacionadas, ya que las distintas partes del sistema están influenciadas por otras partes. Por ejemplo, en la figura PT3.1*a*, el reactor 4 recibe sustancias químicas de los reactores 2 y 3. En consecuencia, su respuesta depende de la cantidad de sustancias químicas en esos reactores.

Cuando esas dependencias se expresan matemáticamente, las ecuaciones resultantes a menudo son de forma algebraica y lineal, como la ecuación (PT3.1). Las x son medidas de las magnitudes de las respuestas de los componentes individuales. Al usar la figura PT3.1*a* como ejemplo, x_1 podría cuantificar la cantidad de masa en el primer reactor, x_2 cuantificaría la cantidad en el segundo, y así sucesivamente. Las a representan comúnmente las propiedades y características relacionadas con las interacciones entre los componentes. Por ejemplo, las a en la figura PT3.1*a* reflejarían las velocidades de masa entre los reactores. Por último, las b representan las funciones forzadas que actúan sobre el sistema, como la velocidad de alimentación en la figura PT3.1*a*. Las aplicaciones en el capítulo 12 proporcionan otros ejemplos de tales ecuaciones obtenidas de la práctica de la ingeniería.

Problemas de multicomponentes de los tipos anteriores surgen tanto de modelos matemáticos de variables *agrupadas* (macro) como *distribuidas* (micro) (figura PT3.1). Los problemas de variables agrupadas involucran componentes finitos relacionadas. Entre los ejemplos se encuentran estructuras (sección 12.2), reactores (figura PT3.1*a* y sección 12.1) y circuitos eléctricos (sección 12.3). Estos tipos de problemas utilizan modelos que ofrecen poco o ningún detalle espacial.

En cambio, los problemas con variables distribuidas intentan describir detalles espaciales de los sistemas sobre una base continua o semicontinua. La distribución de sustancias químicas a lo largo de un reactor rectangular alargado (figura PT3.1*b*) es un ejemplo de un modelo de variable continua. Las ecuaciones diferenciales obtenidas a partir de las leyes de conservación especifican la distribución de la variable dependiente para tales sistemas. Esas ecuaciones diferenciales pueden resolverse numéricamente al convertirlas en un sistema equivalente de ecuaciones algebraicas simultáneas. La solución de tales sistemas de ecuaciones representa una importante área de aplicación a la ingeniería de los métodos en los siguientes capítulos. Esas ecuaciones están relacionadas, ya que las variables en una posición son dependientes de las variables en regiones adyacentes. Por ejemplo, la concentración en la mitad del reactor es una función de la concentración en regiones adyacentes. Ejemplos similares podrían desarrollarse para la distribución espacial de la temperatura o del *momentum*. Más adelante, abordaremos tales problemas cuando analicemos ecuaciones diferenciales.

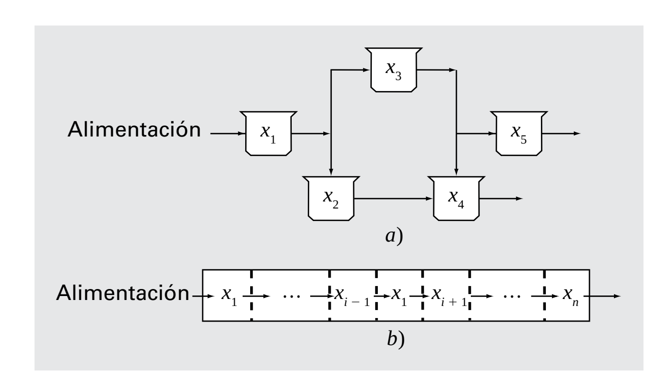


Figura PT3.1
Dos tipos de sistemas que se modelan mediante ecuaciones algebraicas lineales. *a*) sistemas de variables agrupadas que involucran componentes finitos relacionados y *b*) sistemas de variables distribuidas que involucran un continuo.

Además de sistemas físicos, las ecuaciones algebraicas lineales simultáneas surgen también en diferentes contextos de problemas matemáticos. Éstos resultan cuando se requiere de funciones matemáticas que satisfagan varias condiciones en forma simultánea. Cada condición resulta en una ecuación que contiene coefi-

cientes conocidos y variables desconocidas. Las técnicas analizadas en esta parte sirven para encontrar las incógnitas cuando las ecuaciones son lineales y algebraicas. Algunas técnicas numéricas de uso general que emplean ecuaciones simultáneas son el análisis de regresión (capítulo 17) y la interpolación por trazadores (*splines*) (capítulo 18).

PT3.2 ANTECEDENTES MATEMÁTICOS

Todas las partes de este libro requieren de algunos conocimientos matemáticos. Para la parte tres, el álgebra y la notación matricial son útiles, ya que proporcionan una forma concisa para representar y manejar ecuaciones algebraicas lineales. Si usted ya está familiarizado con las matrices, quizá le convenga pasar a la sección PT3.3. Para quienes no tengan un conocimiento previo o necesiten un repaso, el siguiente material ofrece una breve introducción al tema.

PT3.2.1 Notación matricial

Una *matriz* consiste en un arreglo rectangular de elementos representado por un solo símbolo. Como se ilustra en la figura PT3.2, $[A]$ es la notación breve para la matriz y a_{ij} designa un *elemento* individual de la matriz.

Un conjunto horizontal de elementos se llama un *renglón* (o fila); y uno vertical, *columna*. El primer subíndice i siempre designa el número del renglón en el cual está el elemento. El segundo subíndice j designa la columna. Por ejemplo, el elemento a_{23} está en el renglón 2 y la columna 3.

La matriz en la figura PT3.2 tiene n renglones y m columnas, y se dice que tiene una dimensión (o tamaño) de n por m (o $n \times m$). Ésta se conoce como una matriz n por m.

Columna 3   Renglón 2

$$[A] = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1m} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2m} \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ a_{n} & a_{n2} & a_{n3} & \dots & a_{nm} \end{bmatrix}$$

Figura PT3.2
Una matriz.

Matrices con dimensión de renglón $n = 1$, tales como

$$[B] = [b_1 \;\; b_2 \;\; \cdots \;\; b_m]$$

se les conoce como *vectores renglón*. Observe que para simplificar se elimina el primer subíndice de cada elemento. También, debe mencionarse que hay ocasiones en las que se requiere emplear una notación breve especial para distinguir una matriz renglón de otros tipos de matrices. Una forma para llevar a cabo esto es mediante el uso de corchetes abiertos en la parte superior, así $\lfloor B \rfloor$.

Las matrices con dimensión de columna $m = 1$, tales como

$$[C] = \begin{bmatrix} c_1 \\ c_2 \\ \cdot \\ \cdot \\ \cdot \\ c_n \end{bmatrix}$$

se conocen como *vectores columna*. Para simplificar, se elimina el segundo subíndice. Como en el caso del vector renglón, en ocasiones se desea emplear una notación breve especial para distinguir una matriz columna de otros tipos de matrices. Una forma para realizarlo consiste en emplear paréntesis de llave, así $\{C\}$.

A las matrices en las que $n = m$ se les llama *matrices cuadradas*. Por ejemplo, una matriz de 4 por 4 es

$$[A] = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$$

A la diagonal que contiene los elementos a_{11}, a_{22}, a_{33} y a_{44} se le llama *diagonal principal* de la matriz.

Las matrices cuadradas resultan particularmente importantes cuando se resuelven sistemas de ecuaciones lineales simultáneas. En tales sistemas, el número de ecuaciones (que corresponde a los renglones) y el número de incógnitas (que corresponde a las columnas) debe ser igual para que sea posible tener una solución única. En consecuencia, cuando se trabaja con tales sistemas se tienen matrices cuadradas de coeficientes. Algunos tipos especiales de matrices cuadradas se describen en el cuadro PT3.1.

PT3.2.2 Reglas de operaciones con matrices

Ahora que ya especificamos el significado de una matriz, podemos definir algunas reglas de operación que rigen su uso. Dos matrices n por m son iguales si, y sólo si, cada elemento en la primera matriz es igual a cada elemento en la segunda matriz; es decir, $[A] = [B]$ si $a_{ij} = b_{ij}$ para todo i y j.

La *suma* de dos matrices, por ejemplo, $[A]$ y $[B]$, se obtiene al sumar los términos correspondientes de cada matriz. Los elementos de la matriz resultante $[C]$ son:

$$c_{ij} = a_{ij} + b_{ij}$$

para $i = 1, 2, \ldots, n$ y $j = 1, 2, \ldots, m$. De manera similar, la *resta* de dos matrices, por ejemplo, $[E]$ menos $[F]$, se obtiene al restar los términos correspondientes así:

$$d_{ij} = e_{ij} - f_{ij}$$

para $i = 1, 2, \ldots, n$ y $j = 1, 2, \ldots, m$. De las definiciones anteriores se concluye directamente que la suma y la resta sólo pueden realizarse entre matrices que tengan las mismas dimensiones.

Cuadro PT3.1 Tipos especiales de matrices cuadradas

Hay diferentes formas especiales de matrices cuadradas que son importantes y que deben mencionarse:

Una *matriz simétrica* es aquella donde $a_{ij} = a_{ji}$ para todo i y j. Por ejemplo,

$$[A] = \begin{bmatrix} 5 & 1 & 2 \\ 1 & 3 & 7 \\ 2 & 7 & 8 \end{bmatrix}$$

es una matriz simétrica de 3 por 3.

Una *matriz diagonal* es una matriz cuadrada donde todos los elementos fuera de la diagonal principal son iguales a cero,

$$[A] = \begin{bmatrix} a_{11} & & & \\ & a_{22} & & \\ & & a_{33} & \\ & & & a_{44} \end{bmatrix}$$

Observe que donde hay grandes bloques de elementos que son cero, se dejan en blanco.

Una *matriz identidad* es una matriz diagonal donde todos los elementos sobre la diagonal principal son iguales a 1,

$$[I] = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}$$

El símbolo $[I]$ se utiliza para denotar la matriz identidad. La matriz identidad tiene propiedades similares a la unidad.

Una *matriz triangular superior* es aquella donde todos los elementos por debajo de la diagonal principal son cero,

$$[A] = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ & a_{22} & a_{23} & a_{24} \\ & & a_{33} & a_{34} \\ & & & a_{44} \end{bmatrix}$$

Una *matriz triangular inferior* es aquella donde todos los elementos por arriba de la diagonal principal son cero,

$$[A] = \begin{bmatrix} a_{11} & & & \\ a_{21} & a_{22} & & \\ a_{31} & a_{32} & a_{33} & \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$$

Una *matriz a bandas* tiene todos los elementos iguales a cero, con la excepción de una banda centrada sobre la diagonal principal:

$$[A] = \begin{bmatrix} a_{11} & a_{12} & & \\ a_{21} & a_{22} & a_{23} & \\ & a_{32} & a_{33} & a_{34} \\ & & a_{43} & a_{44} \end{bmatrix}$$

La matriz anterior tiene un ancho de banda de 3 y se le da un nombre especial: *matriz tridiagonal*.

Tanto la suma como la resta son *conmutativas*:

$[A] + [B] = [B] + [A]$

La suma y la resta son también *asociativas*, es decir,

$([A] + [B]) + [C] = [A] + ([B] + [C])$

La *multiplicación* de una matriz $[A]$ por un escalar g se obtiene al multiplicar cada elemento de $[A]$ por g, como en

$$[D] = g[A] = \begin{bmatrix} ga_{11} & ga_{12} & \cdots & ga_{1m} \\ ga_{21} & ga_{22} & \cdots & ga_{2m} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ ga_{n1} & ga_{n2} & \cdots & ga_{nm} \end{bmatrix}$$

El producto de dos matrices se representa como $[C] = [A][B]$, donde los elementos de $[C]$ están definidos como (véase cuadro PT3.2 para tener una forma simple de conceptualizar la multiplicación de matrices)

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \qquad \text{(PT3.2)}$$

Cuadro PT3.2 Un método simple para multiplicar dos matrices

Aunque la ecuación (PT3.2) es adecuada para implementarse en una computadora, no es el medio más simple para visualizar la mecánica de multiplicar dos matrices. Lo que sigue es una forma más tangible de entender la operación.

Suponga que queremos multiplicar $[X]$ por $[Y]$ para obtener $[Z]$, donde

$$[Z] = [X][Y] = \begin{bmatrix} 3 & 1 \\ 8 & 6 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} 5 & 9 \\ 7 & 2 \end{bmatrix}$$

Una forma simple para visualizar el cálculo de $[Z]$ es subir $[Y]$ así:

$$\begin{matrix} & & \Uparrow & \\ & & \begin{bmatrix} 5 & 9 \\ 7 & 2 \end{bmatrix} & \leftarrow [Y] \\ [X] \rightarrow & \begin{bmatrix} 3 & 1 \\ 8 & 6 \\ 0 & 4 \end{bmatrix} & \begin{bmatrix} & & \\ & ? & \\ & & \end{bmatrix} & \leftarrow [Z] \end{matrix}$$

Ahora, la matriz $[Z]$ se puede calcular en el espacio dejado por $[Y]$. Este formato es útil, ya que alinea los renglones y columnas apropiados para que se multipliquen. Por ejemplo, de acuerdo con la ecuación (PT3.2), el elemento z_{11} se obtiene al multiplicar el primer renglón de $[X]$ por la primera columna de $[Y]$. Esta cantidad se obtiene al sumar el producto de x_{11} y y_{11} al producto de x_{12} y y_{21} así:

$$\begin{matrix} & \begin{bmatrix} \mathbf{5} & 9 \\ \mathbf{7} & 2 \end{bmatrix} \\ & \downarrow \\ \begin{bmatrix} \mathbf{3} & \mathbf{1} \\ 8 & 6 \\ 0 & 4 \end{bmatrix} \rightarrow & \begin{bmatrix} \mathbf{3 \times 5 + 1 \times 7 = 22} & \\ & \\ & \end{bmatrix} \end{matrix}$$

De esta manera, z_{11} es igual a 22. El elemento z_{21} se calcula de manera semejante así:

$$\begin{matrix} & \begin{bmatrix} \mathbf{5} & 9 \\ \mathbf{7} & 2 \end{bmatrix} \\ & \downarrow \\ \begin{bmatrix} \mathbf{3} & \mathbf{1} \\ 8 & 6 \\ 0 & 4 \end{bmatrix} \rightarrow & \begin{bmatrix} 22 & \\ \mathbf{8 \times 5 + 6 \times 7 = 82} & \\ & \end{bmatrix} \end{matrix}$$

Los cálculos continúan en esta forma, siguiendo la alineación de renglones y columnas, para obtener el resultado

$$[Z] = \begin{bmatrix} 22 & 29 \\ 82 & 84 \\ 28 & 8 \end{bmatrix}$$

Observe cómo este método simple explica el porqué es imposible multiplicar dos matrices si el número de columnas de la primera matriz no es igual al número de renglones en la segunda matriz. Note también la importancia del orden en la multiplicación (es decir, la multiplicación de matrices no es conmutativa).

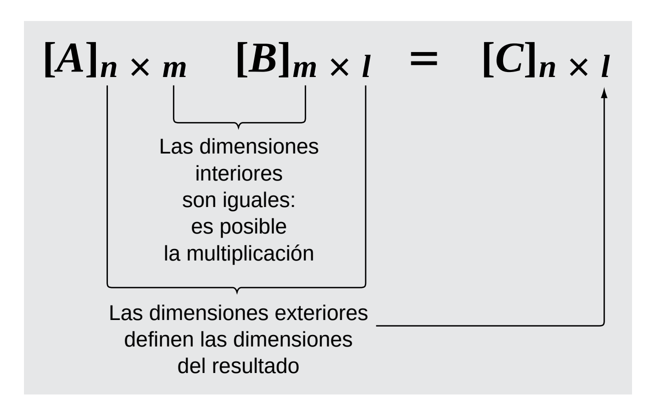


Figura PT3.3

donde n = la dimensión columna de $[A]$ y la dimensión renglón de $[B]$. Es decir, el elemento c_{ij} se obtiene al sumar el producto de elementos individuales del i-ésimo renglón de la primera matriz, en este caso $[A]$, por la j-ésima columna de la segunda matriz $[B]$.

De acuerdo con esta definición, la multiplicación de dos matrices se puede realizar sólo si la primera matriz tiene tantas columnas como el número de renglones en la segunda matriz. Así, si $[A]$ es una matriz n por m, $[B]$ podría ser una matriz m por l. En este caso, la matriz resultante $[C]$ tendrá dimensión n por l. Sin embargo, si $[B]$ fuera una matriz l por m, la multiplicación no podrá ser ejecutada. La figura PT3.3 proporciona una forma fácil para verificar si se pueden multiplicar dos matrices.

Si las dimensiones de las matrices son adecuadas, la multiplicación matricial es *asociativa*,

$$([A][B])[C] = [A]([B][C])$$

y *distributiva*,

$$[A]([B] + [C]) = [A][B] + [A][C]$$

o

$$([A] + [B])[C] = [A][C] + [B][C]$$

Sin embargo, la multiplicación en general no es *conmutativa*:

$$[A][B] \neq [B][A]$$

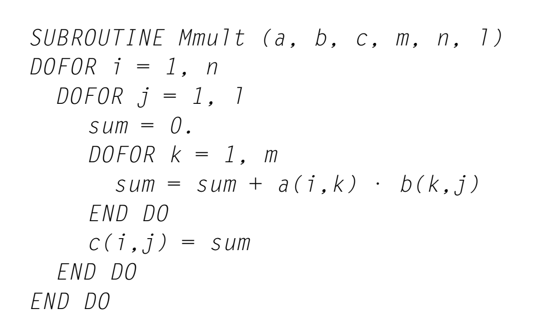
Esto es, el orden de la multiplicación es importante.

La figura PT3.4 muestra el pseudocódigo para multiplicar una matriz $[A]$ n por m, por una matriz $[B]$ m por l, y guardar el resultado en una matriz $[C]$ n por l. Observe que, en lugar de que el producto interno sea directamente acumulado en $[C]$, se recoge en una variable temporal, *sum*. Se hace así por dos razones. Primero, es un poco más eficiente, ya que la computadora necesita determinar la localización de $c_{i,j}$ sólo $n \times l$ veces en lugar de $n \times l \times m$ veces. Segundo, la precisión de la multiplicación puede mejorarse mucho al declarar a *sum* como una variable de doble precisión (recuerde el análisis de productos internos en la sección 3.4.2).

Aunque la multiplicación es posible, la división de matrices no está definida. No obstante, si una matriz $[A]$ es cuadrada y no singular, existe otra matriz $[A]^{-1}$, llamada la *inversa* de $[A]$, para la cual

$$[A][A]^{-1} = [A]^{-1}[A] = [I] \tag{PT3.3}$$

Así, la multiplicación de una matriz por la inversa es análoga a la división, en el sentido de que un número dividido por sí mismo es igual a 1. Es decir, la multiplicación de una matriz por su inversa nos lleva a la matriz identidad (recuerde el cuadro PT3.1).



```
SUBROUTINE Mmult (a, b, c, m, n, l)
DOFOR i = 1, n
  DOFOR j = 1, l
    sum = 0.
    DOFOR k = 1, m
      sum = sum + a(i,k) · b(k,j)
    END DO
    c(i,j) = sum
  END DO
END DO
```

Figura PT3.4

La inversa de una matriz cuadrada bidimensional se representa en forma simple mediante

$$[A]^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \tag{PT3.4}$$

De forma similar, para matrices de dimensiones mayores las fórmulas son más complicadas. Algunas secciones de los capítulos 10 y 11 se dedicarán a técnicas que usen métodos numéricos y la computadora para calcular la inversa de tales sistemas.

Otras dos manipulaciones con matrices que serán útiles para nuestro análisis son la transpuesta y la *traza* de una matriz. La transpuesta de una matriz implica transformar sus renglones en columnas y viceversa. Por ejemplo, dada la matriz de 4×4,

$$[A] = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$$

la transpuesta, designada por $[A]^T$, está definida como

$$[A]^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} & a_{41} \\ a_{12} & a_{22} & a_{32} & a_{42} \\ a_{13} & a_{23} & a_{33} & a_{43} \\ a_{14} & a_{24} & a_{34} & a_{44} \end{bmatrix}$$

En otras palabras, el elemento a_{ij} de la transpuesta es igual al elemento a_{ji} de la matriz original.

La *transpuesta* tiene muchas funciones en álgebra matricial. Una ventaja es que permite escribir un vector columna como un renglón. Por ejemplo, si

$$\{C\} = \begin{Bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{Bmatrix}$$

entonces

$$\{C\}^T = \lfloor c_1 \;\; c_2 \;\; c_3 \;\; c_4 \rfloor$$

donde el superíndice T indica la transpuesta. Por ejemplo, esto puede ahorrar espacio cuando se escribe un vector columna. Además, la transpuesta tiene diversas aplicaciones matemáticas.

La *traza* de una matriz es la suma de los elementos en su diagonal principal, se designa como tr $[A]$ y se calcula como

$$\text{tr}\,[A] = \sum_{i=1}^{n} a_{ii}$$

La traza se usará en el análisis de valores propios en el capítulo 27.

La última manipulación de una matriz que resultará de utilidad para nuestro análisis es el *aumentar.* Una matriz es aumentada al agregar una columna (o columnas) a la matriz original. Por ejemplo, suponga que tenemos una matriz de coeficientes:

$$[A] = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Por ejemplo, se puede aumentar esta matriz $[A]$ con una matriz identidad (recuerde el cuadro PT3.1) para obtener una matriz de dimensiones 3 por 6:

$$[A] = \left[\begin{array}{ccc:ccc} a_{11} & a_{12} & a_{13} & 1 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 & 1 & 0 \\ a_{31} & a_{32} & a_{33} & 0 & 0 & 1 \end{array}\right]$$

Tal expresión es útil cuando debe ejecutarse un conjunto de operaciones idénticas sobre dos matrices. Así, podemos realizar las operaciones sobre una sola matriz aumentada, en lugar de hacerlo sobre dos matrices individuales.

PT3.2.3 Representación de ecuaciones algebraicas lineales en forma matricial

Debe ser claro que las matrices proporcionan una notación concisa para representar ecuaciones lineales simultáneas. Por ejemplo, la ecuación (PT3.1) puede expresarse como

$$[A]\{X\} = \{B\} \tag{PT3.5}$$

donde $[A]$ es la matriz cuadrada n por n de coeficientes,

$$[A] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

$\{B\}$ es el vector columna n por 1 de las constantes,

$$\{B\}^T = \lfloor b_1 \;\; b_2 \;\; \cdots \;\; b_n \rfloor$$

y $\{X\}$ es el vector columna n por 1 de las incógnitas:

$$\{X\}^T = \lfloor x_1 \;\; x_2 \;\; \cdots \;\; x_n \rfloor$$

Recuerde la definición de multiplicación de matrices [ecuación (PT3.2) o cuadro PT3.2] para comprobar que las ecuaciones (PT3.1) y (PT3.5) son equivalentes. También, observe que la ecuación (PT3.5) es una multiplicación matricial válida, ya que el número de columnas, n, de la primera matriz $[A]$, es igual al número de renglones, n, de la segunda matriz $\{X\}$.

Esta parte del libro se dedica a encontrar la solución $\{X\}$ de la ecuación (PT3.5). La manera formal de obtener la solución usando álgebra matricial es multiplicando cada lado de la ecuación por la inversa de $[A]$:

$$[A]^{-1}[A]\{X\} = [A]^{-1}\{B\}$$

Como $[A]^{-1}[A]$ es igual a la matriz identidad, la ecuación se convierte en

$$\{X\} = [A]^{-1}\{B\} \tag{PT3.6}$$

Por lo tanto, se ha encontrado la solución $\{X\}$ de la ecuación. Éste es otro ejemplo de cómo la inversa desempeña un papel importante en el álgebra de matrices que es similar a la división. Debe observarse que ésta no es una forma muy eficiente para resolver un sistema de ecuaciones. Así, se emplean otros procedimientos para construir los algoritmos numéricos. Sin embargo, como se analizó en el capítulo 10, la matriz inversa tiene gran valor en los análisis de ingeniería de tales sistemas.

Por último, algunas veces encontraremos útil aumentar $[A]$ con $\{B\}$. Por ejemplo, si $n = 3$, resultará una matriz de dimensión 3 por 4:

$$[A] = \left[\begin{array}{ccc:c} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_2 \\ a_{31} & a_{32} & a_{33} & b_3 \end{array}\right] \tag{PT3.7}$$

Expresar las ecuaciones en esta forma es útil, ya que varias de las técnicas para resolver sistemas lineales requieren operaciones idénticas en un renglón de coeficientes y en las correspondientes constantes del lado derecho. Como se expresa en la ecuación (PT3.7), es posible realizar las manipulaciones de una vez sobre un renglón de la matriz aumentada, en lugar de hacerlo de manera separada sobre la matriz de coeficientes y en el vector del lado derecho.

PT3.3 Orientación

Antes de presentar los métodos numéricos, será útil una orientación adicional. Lo siguiente pretende ser una visión general del material analizado en la parte tres. Además, se plantean algunos objetivos para ayudarle a enfocar sus esfuerzos al estudiar el material.

PT3.3.1 Alcance y presentación preliminar

La figura PT3.5 proporciona un resumen de la parte tres. El *capítulo 9* se dedica a la técnica fundamental para resolver sistemas algebraicos lineales: la *eliminación de Gauss*. Antes de entrar en un análisis detallado de dicha técnica, una sección preliminar trata de los métodos simples para resolver sistemas pequeños. Esos procedimientos se presentan para ofrecer cierto conocimiento visual y porque uno de los métodos (la eliminación de incógnitas) representa la base para la eliminación de Gauss.

Después del material preliminar, se estudia la eliminación de Gauss "simple". Comenzamos con esta versión "desnuda" debido a que permite elaborar la técnica fundamental sin detalles que la compliquen. Después, en las siguientes secciones, analizamos problemas potenciales del método simple y presentamos diferentes modificaciones para minimizar y evitar tales problemas. Lo esencial en este análisis será el proceso de intercambio de renglones, o *pivoteo parcial*.

El *capítulo 10* empieza ilustrando cómo se puede formular la eliminación de Gauss como una solución por *descomposición LU*. Se trata de técnicas de solución que son valiosas para los casos donde se necesita evaluar muchos vectores del lado derecho. Se muestra cómo este atributo permite

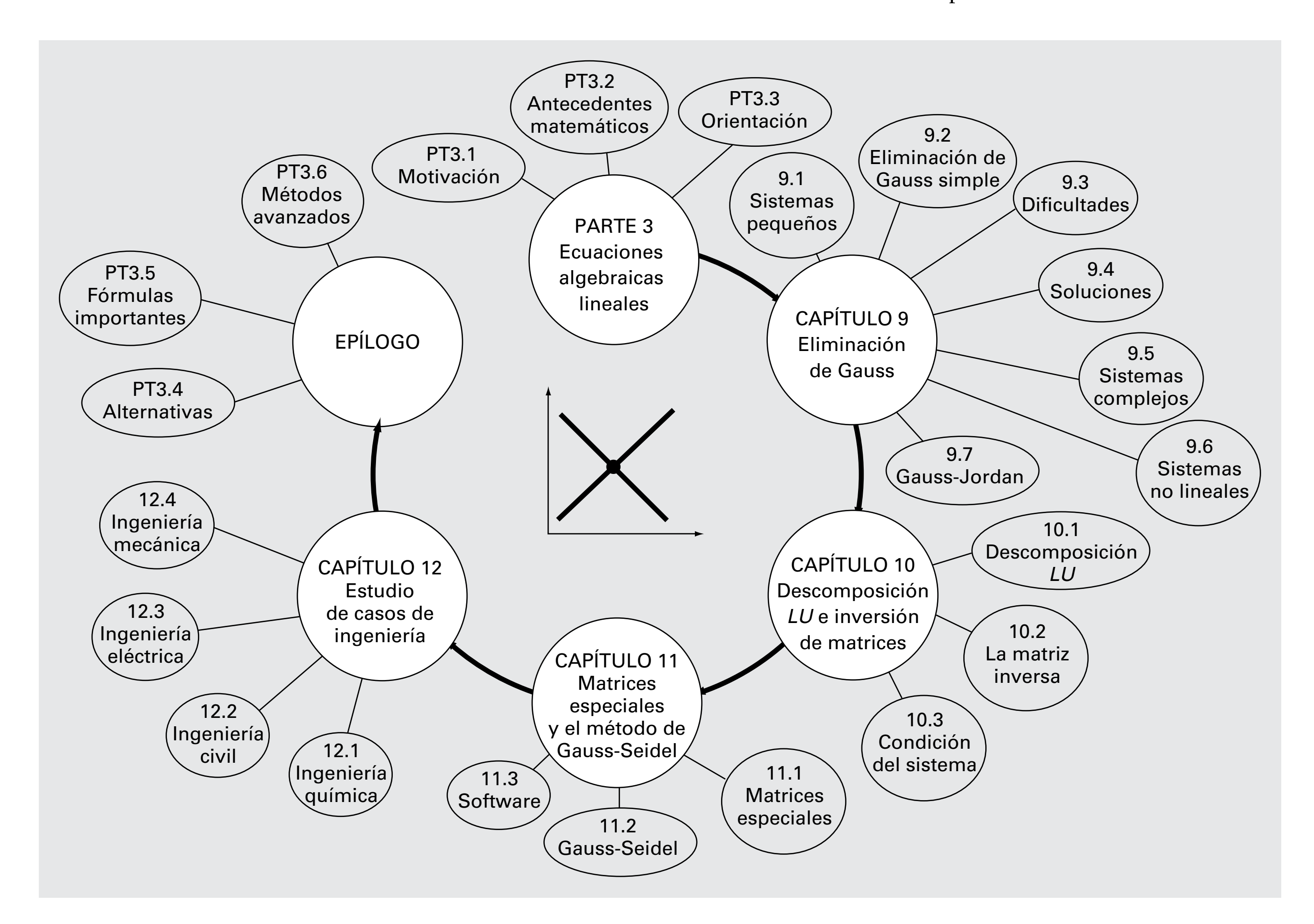


Figura PT3.5
Diagrama esquemático de la organización del material en la parte tres: Ecuaciones algebraicas lineales.

hacer eficiente el cálculo de la *matriz inversa*, la cual tiene una tremenda utilidad en la práctica de la ingeniería. Por último, el capítulo termina con un estudio de la condición matricial. El *número de condición* se presenta como una medida de la pérdida de dígitos significativos de exactitud que puede resultar cuando se resuelven matrices mal condicionadas.

El inicio del *capítulo 11* se concentra en los tipos especiales de sistemas de ecuaciones que tienen una gran aplicación en ingeniería. En particular, se presentan técnicas eficientes para resolver *sistemas tridiagonales*. Después, en el resto del capítulo se centra la atención en una alternativa a los métodos de eliminación llamada el *método de Gauss-Seidel*. Esta técnica es similar en esencia a los métodos aproximados para raíces de ecuaciones que se analizaron en el capítulo 6. Es decir, la técnica consiste en suponer una solución y después iterar para obtener una aproximación mejorada. Al final del capítulo se incluye información relacionada con la solución de ecuaciones algebraicas lineales con ayuda de paquetes de software.

En el *capítulo 12* se muestra cómo se aplican realmente los métodos para la solución de problemas. Como en las otras partes del libro, las aplicaciones se toman de todos los campos de la ingeniería.

Por último, se incluye un epílogo al final de la parte tres. Este repaso comprende un análisis de las ventajas y desventajas relevantes para la implementación de los métodos en la práctica de la ingeniería. Esta sección también resume las fórmulas importantes y los métodos avanzados relacionados con las ecuaciones algebraicas lineales. Como tal, puede usarse antes de los exámenes o en la práctica profesional, a manera de actualización, cuando se tenga que volver a considerar las ecuaciones algebraicas lineales.

PT3.3.2 Metas y objetivos

Objetivos de estudio Al terminar la parte tres, usted será capaz de resolver problemas con ecuaciones algebraicas lineales y de valorar la aplicación de esas ecuaciones en muchos campos de la ingeniería. Deberá esforzarse en dominar varias técnicas y su confiabilidad, así como conocer las ventajas y desventajas para seleccionar el “mejor” método (o métodos) para cualquier problema en particular. Además de estos objetivos generales, deberán asimilarse y dominarse los conceptos específicos enlistados en la tabla PT3.1.

Objetivos de cómputo Sus objetivos de cómputo fundamentales son ser capaz de resolver un sistema de ecuaciones algebraicas lineales y evaluar la matriz inversa. Usted deberá tener subprogramas desarrollados para una descomposición *LU*, tanto de matrices completas como tridiagonales. Quizá desee también tener su propio software para implementar el método Gauss-Seidel.

Deberá saber cómo usar los paquetes para resolver ecuaciones algebraicas lineales y encontrar la matriz inversa. También deberá conocer muy bien la manera en que las mismas evaluaciones se pueden implementar en paquetes de uso común, como Excel, MATLAB y Mathcad.

Tabla PT3.1 Objetivos específicos de estudio de la parte tres.

1. Comprender la interpretación gráfica de sistemas mal condicionados y cómo se relacionan con el determinante.
2. Conocer la terminología: eliminación hacia adelante, sustitución hacia atrás, ecuación pivote y coeficiente pivote.
3. Entender los problemas de división entre cero, errores de redondeo y mal condicionamiento.
4. Saber cómo calcular el determinante con la eliminación de Gauss.
5. Comprender las ventajas del pivoteo; notar la diferencia entre pivoteos parcial y completo.
6. Saber la diferencia fundamental entre el método de eliminación de Gauss y el de Gauss-Jordan y cuál es más eficiente.
7. Reconocer el modo en que la eliminación de Gauss se formula como una descomposición *LU*.
8. Saber cómo incorporar el pivoteo y la inversión de matrices en un algoritmo de descomposición *LU*.
9. Conocer el modo de interpretar los elementos de la matriz inversa al evaluar cálculos de respuesta al estímulo en ingeniería.
10. Percatarse del modo de usar la inversa y las normas de matrices para evaluar la condición de un sistema.
11. Entender cómo los sistemas bandeados y simétricos pueden descomponerse y resolverlos de manera eficiente.
12. Entender por qué el método de Gauss-Seidel es adecuado para grandes sistemas de ecuaciones dispersos.
13. Comprender cómo valorar la diagonal dominante de un sistema de ecuaciones y el modo de relacionarla con el sistema para que pueda resolverse con el método de Gauss-Seidel.
14. Entender la fundamentación de la relajación; saber dónde son apropiadas la bajorrelajación y la sobrerrelajación.

CAPÍTULO 9

Eliminación de Gauss

En este capítulo se analizan las ecuaciones algebraicas lineales simultáneas que en general se representan como

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned} \tag{9. 1}$$

donde las a son los coeficientes constantes y las b son los términos constantes.

La técnica que se describe en este capítulo se conoce como la *eliminación de Gauss*, ya que implica una combinación de ecuaciones para eliminar las incógnitas. Aunque éste es uno de los métodos más antiguos para resolver ecuaciones lineales simultáneas, continúa siendo uno de los algoritmos de mayor importancia, y es la base para resolver ecuaciones lineales en muchos paquetes de software populares.

9.1 SOLUCIÓN DE SISTEMAS PEQUEÑOS DE ECUACIONES

Antes de analizar los métodos computacionales, describiremos algunos métodos que son apropiados en la solución de pequeños sistemas de ecuaciones simultáneas ($n \leq 3$) que no requieren de una computadora. Éstos son el método gráfico, la regla de Cramer y la eliminación de incógnitas.

9.1.1 Método gráfico

Para dos ecuaciones se puede obtener una solución al graficarlas en coordenadas cartesianas con un eje que corresponda a x_1 y el otro a x_2. Debido a que en estos sistemas lineales, cada ecuación se relaciona con una línea recta, lo cual se ilustra fácilmente mediante las ecuaciones generales

$$a_{11}x_1 + a_{12}x_2 = b_1$$

$$a_{21}x_1 + a_{22}x_2 = b_2$$

En ambas ecuaciones se puede despejar x_2:

$$x_2 = -\left(\frac{a_{11}}{a_{12}}\right)x_1 + \frac{b_1}{a_{12}}$$

$$x_2 = -\left(\frac{a_{21}}{a_{22}}\right)x_1 + \frac{b_2}{a_{22}}$$

De esta manera, las ecuaciones ahora están en la forma de líneas rectas; es decir, x_2 = (pendiente) x_1 + intercepto. Tales líneas se grafican en coordenadas cartesianas con x_2 como la ordenada y x_1 como la abscisa. Los valores de x_1 y x_2 en la intersección de las líneas representa la solución.

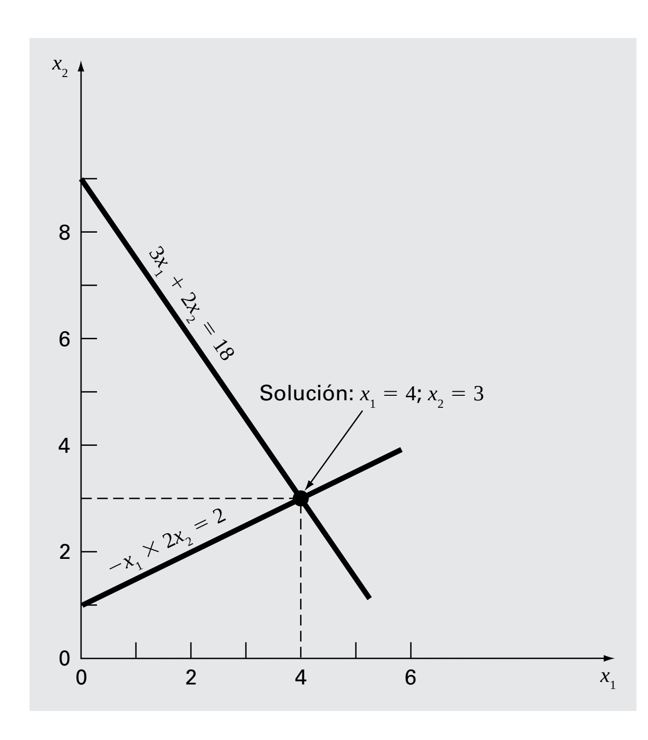


Figura 9.1
Solución gráfica de un conjunto de dos ecuaciones algebraicas lineales simultáneas.
La intersección de las líneas representa la solución.

EJEMPLO 9.1 El método gráfico para dos ecuaciones

Planteamiento del problema Con el método gráfico resuelva

$$3x_1 + 2x_2 = 18 \qquad \text{(E9.1.1)}$$
$$-x_1 + 2x_2 = 2 \qquad \text{(E9.1.2)}$$

Solución Sea x_1 la abscisa. Despejando x_2 de la ecuación (E9.1.1)

$$x_2 = -\frac{3}{2}x_1 + 9$$

la cual, cuando se grafica como en la figura 9.1, es una línea recta con un intercepto en 9 y una pendiente de –3/2.

También de la ecuación (E9.1.2) se despeja x_2:

$$x_2 = \frac{1}{2}x_1 + 1$$

la cual también se grafica en la figura 9.1. La solución es la intersección de las dos líneas en $x_1 = 4$ y $x_2 = 3$. Este resultado se verifica al sustituir los valores en las ecuaciones originales para obtener

$$3(4) + 2(3) = 18$$
$$-(4) + 2(3) = 2$$

De esta manera, los resultados son equivalentes a los valores de la derecha en las ecuaciones originales.

Para tres ecuaciones simultáneas, cada ecuación se representa como un plano en un sistema de coordenadas tridimensional. El punto en donde se intersecan los tres planos representa la solución. Para más de tres incógnitas, los métodos gráficos no funcionan y, por consiguiente, tienen poco valor práctico para resolver ecuaciones simultáneas. No obstante, resultan útiles para visualizar propiedades de las soluciones. Por ejemplo, la figura 9.2 muestra tres casos que pueden ocasionar problemas al resolver sistemas de ecuaciones lineales. La figura 9.2*a* presenta el caso en que las dos ecuaciones representan líneas paralelas. En estos casos no existe solución, ya que las dos líneas jamás se cruzan.

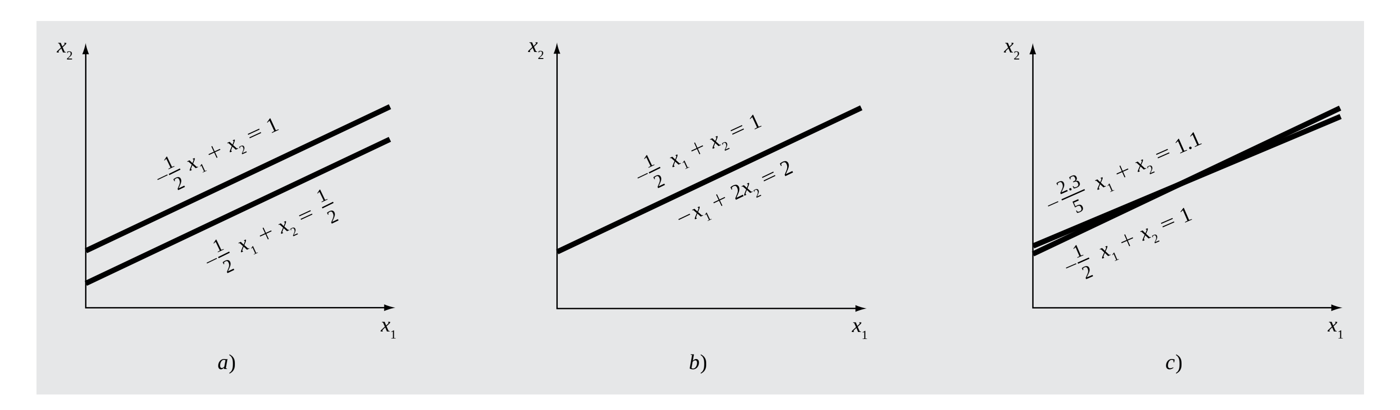


Figura 9.2
Representación gráfica de sistemas singulares y mal condicionados: *a*) no hay solución, *b*) hay una infinidad de soluciones y *c*) sistema mal condicionado donde las pendientes son tan cercanas que es difícil detectar visualmente el punto de intersección.

La figura 9.2*b* representa el caso en que las dos líneas coinciden. En éste existe un número infinito de soluciones. Se dice que ambos tipos de sistemas son *singulares*. Además, los sistemas muy próximos a ser singulares (figura 9.2*c*) también pueden causar problemas; a estos sistemas se les llama *mal condicionados*. Gráficamente, esto corresponde al hecho de que resulta difícil identificar el punto exacto donde las líneas se intersecan. Los sistemas mal condicionados presentan problemas cuando se encuentran durante la solución numérica de ecuaciones lineales, lo cual se debe a que este tipo de sistemas son extremadamente sensibles a los errores de redondeo (recuerde la sección 4.2.3).

9.1.2 Determinantes y la regla de Cramer

La regla de Cramer es otra técnica de solución adecuada para un sistema pequeño de ecuaciones. Antes de hacer una descripción de tal método, se mencionará en forma breve el concepto de determinante que se utiliza en la regla de Cramer. Además, el determinante tiene relevancia en la evaluación del mal condicionamiento de una matriz.

Determinantes El determinante se puede ilustrar para un sistema de tres ecuaciones simultáneas:

$$[A]\{X\} = \{B\}$$

donde $[A]$ es la matriz de coeficientes:

$$[A] = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

El determinante D de este sistema se forma, a partir de los coeficientes de la ecuación, de la siguiente manera:

$$D = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \tag{9.2}$$

Aunque el determinante D y la matriz de coeficientes $[A]$ se componen de los mismos elementos, son conceptos matemáticos completamente diferentes. Por esto, para distinguirlos visualmente se emplean corchetes para encerrar la matriz y líneas rectas verticales para el determinante. En contraste con una matriz, el determinante es un simple número. Por ejemplo, el valor del determinante de segundo orden

$$D = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

se calcula como

$$D = a_{11}a_{22} - a_{12}a_{21} \tag{9.3}$$

En el caso del determinante de tercer orden [ecuación (9.2)], el determinante, que es un simple valor numérico, se calcula así

$$D = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \tag{9.4}$$

donde a los determinantes de 2 por 2 se les llama ***menores***.

EJEMPLO 9.2 Determinantes

Planteamiento del problema Calcule los valores para los determinantes de los sistemas representados en las figuras 9.1 y 9.2.

Solución Para la figura 9.1:

$$D = \begin{vmatrix} 3 & 2 \\ -1 & 2 \end{vmatrix} = 3(2) - 2(-1) = 8$$

Para la figura 9.2*a*:

$$D = \begin{vmatrix} -1/2 & 1 \\ -1/2 & 1 \end{vmatrix} = \frac{-1}{2}(1) - 1\left(\frac{-1}{2}\right) = 0$$

Para la figura 9.2*b*:

$$D = \begin{vmatrix} -1/2 & 1 \\ -1 & 2 \end{vmatrix} = \frac{-1}{2}(2) - 1(-1) = 0$$

Para la figura 9.2*c*:

$$D = \begin{vmatrix} -1/2 & 1 \\ -2.3/5 & 1 \end{vmatrix} = \frac{-1}{2}(1) - 1\left(\frac{-2.3}{5}\right) = -0.04$$

En el ejemplo anterior, los sistemas singulares tienen determinante cero. Además, los resultados sugieren que el sistema que sea casi singular (figura 9.2*c*) tiene un determinante cercano a cero. Estas ideas se tratarán también en análisis subsecuentes de mal condicionamiento (sección 9.3.3).

Regla de Cramer Esta regla establece que cada incógnita de un sistema de ecuaciones lineales algebraicas puede expresarse como una fracción de dos determinantes con denominador D y con el numerador obtenido a partir de D, al reemplazar la columna de coeficientes de la incógnita en cuestión por las constantes $b_1, b_2, \ldots, b_n$. Por ejemplo, x_1 se calcula como

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{vmatrix}}{D} \tag{9.5}$$

EJEMPLO 9.3 Regla de Cramer

Planteamiento del problema Utilice la regla de Cramer para resolver

$$0.3x_1 + 0.52x_2 + x_3 = -0.01$$
$$0.5x_1 + x_2 + 1.9x_3 = 0.67$$
$$0.1x_1 + 0.3x_2 + 0.5x_3 = -0.44$$

Solución El determinante D se puede escribir como [ecuación (9.2)]

$$D = \begin{vmatrix} 0.3 & 0.52 & 1 \\ 0.5 & 1 & 1.9 \\ 0.1 & 0.3 & 0.5 \end{vmatrix}$$

Los menores son [ecuación (9.3)]

$$A_1 = \begin{vmatrix} 1 & 1.9 \\ 0.3 & 0.5 \end{vmatrix} = 1(0.5) - 1.9(0.3) = -0.07$$

$$A_2 = \begin{vmatrix} 0.5 & 1.9 \\ 0.1 & 0.5 \end{vmatrix} = 0.5(0.5) - 1.9(0.1) = 0.06$$

$$A_3 = \begin{vmatrix} 0.5 & 1 \\ 0.1 & 0.3 \end{vmatrix} = 0.5(0.3) - 1(0.1) = 0.05$$

Éstos se usan para evaluar el determinante, como en [ecuación (9.4)]

$$D = 0.3(-0.07) - 0.52(0.06) + 1(0.05) = -0.0022$$

Aplicando la ecuación (9.5), la solución es

$$x_1 = \frac{\begin{vmatrix} -0.01 & 0.52 & 1 \\ 0.67 & 1 & 1.9 \\ -0.44 & 0.3 & 0.5 \end{vmatrix}}{-0.0022} = \frac{0.03278}{-0.0022} = -14.9$$

$$x_2 = \frac{\begin{vmatrix} 0.3 & -0.01 & 1 \\ 0.5 & 0.67 & 1.9 \\ 0.1 & -0.44 & 0.5 \end{vmatrix}}{-0.0022} = \frac{0.0649}{-0.0022} = -29.5$$

$$x_3 = \frac{\begin{vmatrix} 0.3 & 0.52 & -0.01 \\ 0.5 & 1 & 0.67 \\ 0.1 & 0.3 & -0.44 \end{vmatrix}}{-0.0022} = \frac{-0.04356}{-0.0022} = 19.8$$

Para más de tres ecuaciones, la regla de Cramer no resulta práctica, ya que, conforme aumenta el número de ecuaciones, los determinantes consumen tiempo al evaluarlos manualmente (o por computadora). Por consiguiente, se usan otras alternativas más eficientes. Algunas de éstas se basan en la última técnica, sin el uso de la computadora, que se analizará en la siguiente sección: la eliminación de incógnitas.

9.1.3 La eliminación de incógnitas

La eliminación de incógnitas mediante la combinación de ecuaciones es un método algebraico que se ilustra con un sistema de dos ecuaciones simultáneas:

$$a_{11}x_1 + a_{12}x_2 = b_1 \tag{9.6}$$

$$a_{21}x_1 + a_{22}x_2 = b_2 \tag{9.7}$$

La estrategia básica consiste en multiplicar las ecuaciones por constantes, de tal forma que se elimine una de las incógnitas cuando se combinen las dos ecuaciones. El resultado es una sola ecuación en la que se puede despejar la incógnita restante. Este valor se sustituye en cualquiera de las ecuaciones originales para calcular la otra variable.

Por ejemplo, la ecuación (9.6) se multiplica por a_{21} y la ecuación (9.7) por a_{11} para dar

$$a_{11}a_{21}x_1 + a_{12}a_{21}x_2 = b_1a_{21} \tag{9.8}$$

$$a_{21}a_{11}x_1 + a_{22}a_{11}x_2 = b_2a_{11} \tag{9.9}$$

Restando la ecuación (9.8) de la (9.9) se elimina el término x_1 de las ecuaciones para obtener

$$a_{22}a_{11}x_2 - a_{12}a_{21}x_2 = b_2a_{11} - b_1a_{21}$$

de donde se puede despejar

$$x_2 = \frac{a_{11}b_2 - a_{21}b_1}{a_{11}a_{22} - a_{12}a_{21}} \tag{9.10}$$

La ecuación (9.10) se puede entonces sustituir en la ecuación (9.6), de la cual se puede despejar

$$x_1 = \frac{a_{22}b_1 - a_{12}b_2}{a_{11}a_{22} - a_{12}a_{21}} \tag{9.11}$$

Observe que las ecuaciones (9.10) y (9.11) se relacionan directamente con la regla de Cramer, que establece

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} = \frac{b_1a_{22} - a_{12}b_2}{a_{11}a_{22} - a_{12}a_{21}}$$

$$x_2 = \frac{\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} = \frac{a_{11}b_2 - b_1a_{21}}{a_{11}a_{22} - a_{12}a_{21}}$$

EJEMPLO 9.4 Eliminación de incógnitas

Planteamiento del problema Use la eliminación de incógnitas para resolver (recuerde el ejemplo 9.1)

$$3x_1 + 2x_2 = 18$$
$$-x_1 + 2x_2 = 2$$

Solución Utilizando las ecuaciones (9.11) y (9.10),

$$x_1 = \frac{2(18) - 2(2)}{3(2) - 2(-1)} = 4$$

$$x_2 = \frac{3(2) - (-1)18}{3(2) - 2(-1)} = 3$$

cuyos valores coinciden con la solución gráfica (figura 9.1).

La eliminación de incógnitas se puede extender a sistemas con más de dos o tres ecuaciones. Sin embargo, los múltiples cálculos que se requieren para sistemas más grandes hacen que el método sea extremadamente tedioso para realizarse a mano. No obstante, como se describe en la siguiente sección, la técnica llega a formalizarse y programarse fácilmente en la computadora.

9.2 ELIMINACIÓN DE GAUSS SIMPLE

En la sección anterior se utilizó la eliminación de incógnitas para resolver un par de ecuaciones simultáneas. El procedimiento consistió de dos pasos:

1. Las ecuaciones se manipularon para eliminar una de las incógnitas de las ecuaciones. El resultado de este paso de *eliminación* fue el de una sola ecuación con una incógnita.

2. En consecuencia, esta ecuación se pudo resolver directamente y el resultado *sustituirse atrás* en una de las ecuaciones originales para encontrar la incógnita restante.

Esta técnica básica puede extenderse a sistemas grandes de ecuaciones desarrollando un esquema sistemático o algorítmico para eliminar incógnitas y sustituir hacia atrás. La *eliminación de Gauss* es el más básico de dichos esquemas.

Esta sección presenta las técnicas sistemáticas para la eliminación hacia adelante y la sustitución hacia atrás que la eliminación gaussiana comprende. Aunque tales técnicas son muy adecuadas para utilizarlas en computadoras, se requiere de algunas modificaciones para obtener un algoritmo confiable. En particular, el programa debe evitar la división entre cero. Al método siguiente se le llama *eliminación gaussiana "simple"*, ya que no evita este problema. En las siguientes secciones se verán algunas características adicionales necesarias para obtener un programa de cómputo efectivo.

El método está ideado para resolver un sistema general de n ecuaciones:

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = b_1 \tag{9.12a}$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n = b_2 \tag{9.12b}$$

$$\vdots \qquad \qquad \vdots$$

$$a_{n1}x_1 + a_{n2}x_2 + a_{n3}x_3 + \cdots + a_{nn}x_n = b_n \tag{9.12c}$$

Como en el caso de dos ecuaciones, la técnica para resolver n ecuaciones consiste en dos fases: la eliminación de las incógnitas y su solución mediante sustitución hacia atrás.

Eliminación hacia adelante de incógnitas La primera fase consiste en reducir el conjunto de ecuaciones a un sistema triangular superior (figura 9.3). El paso inicial será eliminar la primera incógnita, x_1, desde la segunda hasta la n-ésima ecuación. Para ello, se multiplica la ecuación (9.12a) por a_{21}/a_{11} para obtener

$$a_{21}x_1 + \frac{a_{21}}{a_{11}}a_{12}x_2 + \cdots + \frac{a_{21}}{a_{11}}a_{1n}x_n = \frac{a_{21}}{a_{11}}b_1 \tag{9.13}$$

Ahora, esta ecuación se resta de la ecuación (9.12b) para dar

$$\left(a_{22} - \frac{a_{21}}{a_{11}}a_{12}\right)x_2 + \cdots + \left(a_{2n} - \frac{a_{21}}{a_{11}}a_{1n}\right)x_n = b_2 - \frac{a_{21}}{a_{11}}b_1$$

o

$$a'_{22}x_2 + \cdots + a'_{2n}x_n = b'_2$$

donde el superíndice prima indica que los elementos han cambiado sus valores originales.

El procedimiento se repite después con las ecuaciones restantes. Por ejemplo, la ecuación (9.12a) se puede multiplicar por a_{31}/a_{11} y el resultado se resta de la tercera ecuación. Se repite el procedimiento con las ecuaciones restantes y da como resultado el siguiente sistema modificado:

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = b_1 \tag{9.14a}$$

$$a'_{22}x_2 + a'_{23}x_3 + \cdots + a'_{2n}x_n = b'_2 \tag{9.14b}$$

$$a'_{32}x_2 + a'_{33}x_3 + \cdots + a'_{3n}x_n = b'_3 \tag{9.14c}$$

$$\vdots \qquad \qquad \vdots$$

$$a'_{n2}x_2 + a'_{n3}x_3 + \cdots + a'_{nn}x_n = b'_n \tag{9.14d}$$

En los pasos anteriores, la ecuación (9.12a) se llama la *ecuación pivote*, y a_{11} se denomina el *coeficiente* o *elemento pivote*. Observe que el proceso de

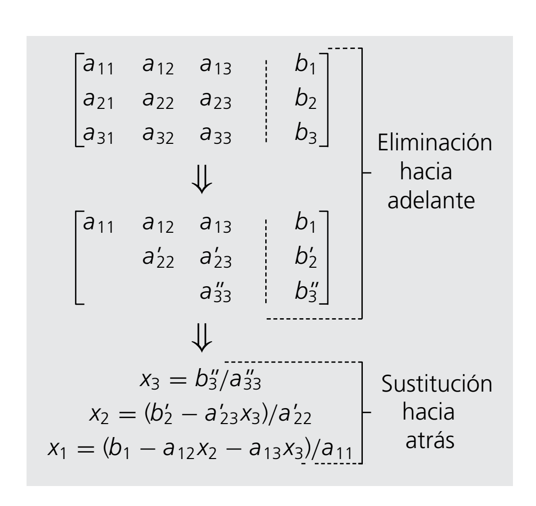


Figura 9.3
Las dos fases de la eliminación de Gauss: eliminación hacia adelante y sustitución hacia atrás. Los superíndices prima indican el número de veces que se han modificado los coeficientes y constantes.

multiplicación del primer renglón por a_{21}/a_{11} es equivalente a dividirla entre a_{11} y multiplicarla por a_{21}. Algunas veces la operación de división es referida como normalización. Se hace esta distinción porque un elemento pivote cero llega a interferir con la normalización al causar una división entre cero. Más adelante se regresará a este punto importante, una vez que se complete la descripción de la eliminación de Gauss simple.

Ahora se repite el procedimiento antes descrito para eliminar la segunda incógnita en las ecuaciones (9.14*c*) hasta (9.14*d*). Para realizar esto, multiplique la ecuación (9.14*b*) por a'_{32}/a'_{22} y reste el resultado de la ecuación (9.14*c*). Se realiza la eliminación en forma similar en las ecuaciones restantes para obtener

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= b_1 \\
a'_{22}x_2 + a'_{23}x_3 + \cdots + a'_{2n}x_n &= b'_2 \\
a''_{33}x_3 + \cdots + a''_{3n}x_n &= b''_3 \\
\vdots \qquad\qquad & \vdots \\
a''_{n3}x_3 + \cdots + a''_{nn}x_n &= b''_n
\end{aligned}
$$

donde el superíndice biprima indica que los elementos se han modificado dos veces.

El procedimiento puede continuar usando las ecuaciones pivote restantes. La última manipulación en esta secuencia es el uso de la $(n - 1)$-ésima ecuación para eliminar el término x_{n-1} de la n-ésima ecuación. Aquí el sistema se habrá transformado en un sistema triangular superior (véase el cuadro PT3.1):

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = b_1 \tag{9.15a}$$

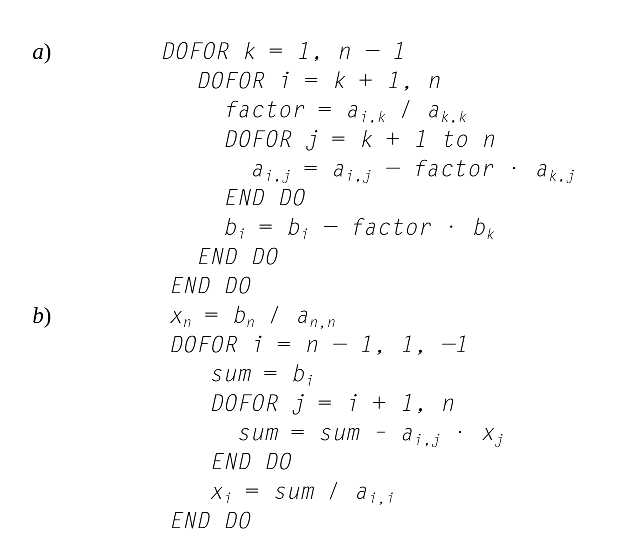
$$a'_{22}x_2 + a'_{23}x_3 + \cdots + a'_{2n}x_n = b'_2 \tag{9.15b}$$

$$a''_{33}x_3 + \cdots + a''_{3n}x_n = b''_3 \tag{9.15c}$$

$$\vdots \qquad\qquad \vdots$$

$$a_{nn}^{(n-1)}x_n = b_n^{(n-1)} \tag{9.15d}$$

El pseudocódigo para implementar la eliminación hacia adelante se presenta en la figura 9.4*a*. Observe que tres *loops* o ciclos anidados proporcionan una representación concisa del proceso. El ciclo externo mueve hacia abajo de la matriz el renglón pivote. El siguiente ciclo mueve hacia abajo el renglón pivote a cada renglón subsecuente, donde la eliminación se llevará a cabo. Finalmente, el ciclo más interno avanza a través de las columnas para eliminar o transformar los elementos de un renglón determinado.



```
a)  DOFOR k = 1, n − 1
      DOFOR i = k + 1, n
        factor = a(i,k) / a(k,k)
        DOFOR j = k + 1 to n
          a(i,j) = a(i,j) − factor · a(k,j)
        END DO
        b(i) = b(i) − factor · b(k)
      END DO
    END DO

b)  x(n) = b(n) / a(n,n)
    DOFOR i = n − 1, 1, −1
      sum = b(i)
      DOFOR j = i + 1, n
        sum = sum − a(i,j) · x(j)
      END DO
      x(i) = sum / a(i,i)
    END DO
```

Figura 9.4
Pseudocódigo que realiza *a*) la eliminación hacia adelante y *b*) la sustitución hacia atrás.

Sustitución hacia atrás De la ecuación (9.15*d*) ahora se despeja x_n:

$$x_n = \frac{b_n^{(n-1)}}{a_{nn}^{(n-1)}} \tag{9.16}$$

Este resultado se puede sustituir hacia atrás en la $(n - 1)$-ésima ecuación y despejar x_{n-1}. El procedimiento, que se repite para evaluar las x restantes, se representa mediante la fórmula:

$$x_i = \frac{b_i^{(i-1)} - \sum_{j=i+1}^{n} a_{ij}^{(i-1)} x_j}{a_{ii}^{(i-1)}} \qquad \text{para } i = n-1, n-2, \ldots, 1 \tag{9.17}$$

El pseudocódigo para implementar las ecuaciones (9.16) y (9.17) se representa en la figura 9.4*b*. Observe la similitud entre este pseudocódigo y el mostrado en la figura PT3.4 para la multiplicación de matrices. De la misma forma que en la figura PT3.4, se utiliza una variable temporal *sum* para acumular la sumatoria de la ecuación (9.17). Esto da por resultado un tiempo de ejecución más rápido que si la sumatoria fuera acumulada en b_i. Más importante aún es que esto permite una mayor eficiencia en la precisión si la variable, *sum*, se declara como variable de doble precisión.

EJEMPLO 9.5 Eliminación de Gauss simple

Planteamiento del problema Emplee la eliminación de Gauss para resolver

$$3x_1 - 0.1x_2 - 0.2x_3 = 7.85 \quad \text{(E9.5.1)}$$

$$0.1x_1 + 7x_2 - 0.3x_3 = -19.3 \quad \text{(E9.5.2)}$$

$$0.3x_1 - 0.2x_2 + 10x_3 = 71.4 \quad \text{(E9.5.3)}$$

Efectúe los cálculos con seis cifras significativas.

Solución La primera parte del procedimiento es la eliminación hacia adelante. Se multiplica la ecuación (E9.5.1) por (0.1)/3 y se resta el resultado de la ecuación (E9.5.2) para obtener

$$7.00333x_2 - 0.293333x_3 = -19.5617$$

Después, se multiplica la ecuación (E9.5.1) por (0.3)/3 y se resta de la ecuación (E9.5.3) para eliminar x_1. Luego de efectuar estas operaciones, el sistema de ecuaciones es

$$3x_1 \quad -0.1x_2 \quad -0.2x_3 = 7.85 \quad \text{(E9.5.4)}$$

$$7.00333x_2 - 0.293333x_3 = -19.5617 \quad \text{(E9.5.5)}$$

$$-0.190000x_2 + 10.0200x_3 = 70.6150 \quad \text{(E9.5.6)}$$

Para completar la eliminación hacia adelante, x_2 debe eliminarse de la ecuación (E9.5.6). Para llevar a cabo esto, se multiplica la ecuación (E9.5.5) por –0.190000/7.00333 y se resta el resultado de la ecuación (E9.5.6). Esto elimina x_2 de la tercera ecuación y reduce el sistema a una forma triangular superior:

$$3x_1 \quad -0.1x_2 \quad -0.2x_3 = 7.85 \quad \text{(E9.5.7)}$$

$$7.00333x_2 - 0.293333x_3 = -19.5617 \quad \text{(E9.5.8)}$$

$$10.0120x_3 = 70.0843 \quad \text{(E9.5.9)}$$

Ahora se pueden resolver estas ecuaciones por sustitución hacia atrás. En primer lugar, de la ecuación (E9.5.9) se despeja x_3

$$x_3 = \frac{70.0843}{10.0120} = 7.0000 \quad \text{(E9.5.10)}$$

Este resultado se sustituye en la ecuación (E9.5.8):

$$7.00333x_2 - 0.293333(7.0000) = -19.5617$$

de la que se despeja

$$x_2 = \frac{-19.5617 + 0.293333(7.0000)}{7.00333} = -2.50000 \quad \text{(E9.5.11)}$$

Por último, las ecuaciones (E9.5.10) y (E9.5.11) se sustituyen en la (E9.5.4):

$$3x_1 - 0.1(-2.50000) - 0.2(7.0000) = 7.85$$

de donde se puede despejar

$$x_1 = \frac{7.85 + 0.1(-2.50000) + 0.2(7.0000)}{3} = 3.00000$$

Los resultados son muy cercanos a la solución exacta, $x_1 = 3$, $x_2 = -2.5$ y $x_3 = 7$. Esto se puede verificar sustituyendo los resultados en el sistema de ecuaciones original:

$$3(3) - 0.1(-2.5) - 0.2(7) = 7.85$$
$$0.1(3) + 7(-2.5) - 0.3(7) = -19.3$$
$$0.3(3) - 0.2(-2.5) + 10(7) = 71.4$$

9.2.1 Conteo de las operaciones

El tiempo de ejecución de la eliminación de Gauss depende de la cantidad de *operaciones con punto flotante* (*floating point operations* = "*flops*") incluidas en el algoritmo. En las computadoras modernas que usan coprocesadores matemáticos, el tiempo empleado para realizar suma/resta y multiplicación/división es aproximadamente el mismo. Por lo tanto, la totalización de estas operaciones permite comprender cuáles partes del algoritmo consumen más tiempo y cómo aumenta el tiempo de computación al hacerse más grande el sistema.

Antes de analizar la eliminación de Gauss simple, primero se definirán algunas cantidades que facilitan el conteo de operaciones:

$$\sum_{i=1}^{m} cf(i) = c\sum_{i=1}^{m} f(i) \qquad \sum_{i=1}^{m} f(i) + g(i) = \sum_{i=1}^{m} f(i) + \sum_{i=1}^{m} g(i) \tag{9.18a, b}$$

$$\sum_{i=1}^{m} 1 = 1 + 1 + 1 + \cdots + 1 = m \qquad \sum_{i=k}^{m} 1 = m - k + 1 \tag{9.18c, d}$$

$$\sum_{i=1}^{m} i = 1 + 2 + 3 + \cdots + m = \frac{m(m+1)}{2} = \frac{m^2}{2} + O(m) \tag{9.18e}$$

$$\sum_{i=1}^{m} i^2 = 1^2 + 2^2 + 3^2 + \cdots + m^2 = \frac{m(m+1)(2m+1)}{6} = \frac{m^3}{3} + O(m^2) \tag{9.18f}$$

donde $O(m^n)$ significa "términos de orden m^n y menores".

Ahora se examinará en forma detallada el algoritmo de la eliminación de Gauss simple (figura 9.4). Primero se contarán los flops en la etapa de la eliminación. En el primer paso durante el *loop* o ciclo externo, $k = 1$. Por lo tanto, los límites del ciclo intermedio son desde $i = 2$ hasta n. De acuerdo con la ecuación (9.18*d*), esto significa que el número de iteraciones en el ciclo intermedio será

$$\sum_{i=2}^{n} 1 = n - 2 + 1 = n - 1 \tag{9.19}$$

Ahora, para cada una de estas iteraciones, hay una división para definir el factor. El ciclo interno realiza después una sola multiplicación y resta para cada iteración de $j = 2$ a n. Por último, hay una multiplicación y resta adicional del valor del lado derecho. Así, en cada iteración del ciclo intermedio, el número de multiplicaciones es

$$1 + [n - 2 + 1] + 1 = 1 + n \tag{9.20}$$

Las multiplicaciones totales para el primer paso a través del ciclo exterior se obtienen, por lo tanto, multiplicando la ecuación (9.19) por la (9.20), para dar $[n - 1](1 + n)$. De modo similar, el número de restas se calcula como $[n - 1](n)$.

Se puede usar un razonamiento similar para estimar los flops para las iteraciones subsecuentes del ciclo externo. Esto se puede resumir como se muestra en la tabla de la página siguiente.

Por lo tanto, el total de flops de la suma/resta para el proceso de eliminación se calcula como

$$\sum_{k=1}^{n-1} (n-k)(n+1-k) = \sum_{k=1}^{n-1} [n(n+1) - k(2n+1) + k^2]$$

Ciclo externo (*k*)	Ciclo medio (*i*)	Flops de suma/resta	Flops de multiplicación/división
1	2, *n*	$(n-1)(n)$	$(n-1)(n+1)$
2	3, *n*	$(n-2)(n-1)$	$(n-2)(n)$
.	.		
.	.		
.	.		
k	*k* + 1, n	$(n-k)(n+1-k)$	$(n-k)(n+2-k)$
.	.		
.	.		
.	.		
n – 1	*n*, *n*	(1) (2)	(1) (3)

o bien

$$n(n+1)\sum_{k=1}^{n-1}1-(2n+1)\sum_{k=1}^{n-1}k+\sum_{k=1}^{n-1}k^2$$

Al aplicar alguna de las relaciones de la ecuación (9.18) se obtiene

$$[n^3+O(n)]-[n^3+O(n^2)+\left[\frac{1}{3}n^3+O(n^2)\right]=\frac{n^3}{3}+O(n) \tag{9.21}$$

Un análisis similar para los flops de la multiplicación/división lleva a lo siguiente

$$[n^3+O(n^2)]-[n^3+O(n)+\left[\frac{1}{3}n^3+O(n^2)\right]=\frac{n^3}{3}+O(n^2) \tag{9.22}$$

Al sumar el resultado queda

$$\frac{2n^3}{3}+O(n^2)$$

Así, el número total de flops es igual a $2n^3/3$ más un componente adicional de proporcionalidad para términos de orden n^2 y menores. El resultado se escribe de esta manera porque conforme n crece, los términos $O(n^2)$ y menores se hacen despreciables. Por lo tanto, se justifica concluir que para un valor de n grande, el esfuerzo necesario para la eliminación hacia adelante converge a $2n^3/3$.

Debido a que sólo se utiliza un *loop* o ciclo, la sustitución hacia atrás es mucho más fácil de evaluar. El número de flops adicionales para la suma/resta es igual a $n(n-1)/2$. Debido a la división adicional anterior al ciclo, el número de flops para la multiplicación/división es $n(n+1)/2$. Esto se suma para llegar a un total de

$$n^2+O(n)$$

Entonces, el trabajo total en la eliminación de Gauss simple se representa como

$$\underbrace{\frac{2n^3}{3}+O(n^2)}_{\text{Eliminación hacia adelante}}+\underbrace{n^2+O(n)}_{\text{Sustitución hacia atrás}}\xrightarrow[\text{aumenta}]{\text{conforme } n}\frac{2n^3}{3}+O(n^2) \tag{9.23}$$

En este análisis destacan dos conclusiones generales útiles:

1. Conforme el sistema se vuelve más grande, el tiempo de cálculo aumenta enormemente. Como en la tabla 9.1, la cantidad de flops aumenta casi tres órdenes de magnitud por cada orden de aumento de la dimensión.
2. La mayor parte del trabajo ocurre en el paso de eliminación. Así, para hacer el método más eficiente, debería enfocarse a este paso.

Tabla 9.1 Número de flops en la eliminación de Gauss simple.

n	Eliminación	Sustitución hacia atrás	Flops totales	$2n^3/3$	Porcentaje debido a la eliminación
10	705	100	805	667	87.58%
100	671 550	10 000	681 550	666 667	98.53%
1 000	6.67×10^8	1×10^6	6.68×10^8	6.67×10^8	99.85%

9.3 DIFICULTADES EN LOS MÉTODOS DE ELIMINACIÓN

Mientras que hay muchos sistemas de ecuaciones que se pueden resolver con la eliminación de Gauss simple, existen algunas dificultades que se deben analizar, antes de escribir un programa de cómputo general donde se implemente el método. Aunque el siguiente material se relaciona en forma directa con la eliminación de Gauss simple, la información también es relevante para otras técnicas de eliminación.

9.3.1 División entre cero

La razón principal por la que se le ha llamado *simple* al método anterior se debe a que durante las fases de eliminación y sustitución hacia atrás es posible que ocurra una división entre cero. Por ejemplo, si se utiliza el método de eliminación de Gauss simple para resolver

$$
\begin{aligned}
2x_2 + 3x_3 &= 8 \\
4x_1 + 6x_2 + 7x_3 &= -3 \\
2x_1 + x_2 + 6x_3 &= 5
\end{aligned}
$$

en la normalización del primer renglón habrá una división entre $a_{11} = 0$. También se pueden presentar problemas cuando un coeficiente está muy cercano a cero. La técnica de *pivoteo* se ha desarrollado para evitar en forma parcial estos problemas. Ésta se describe en la sección 9.4.2.

9.3.2 Errores de redondeo

Aun cuando la solución del ejemplo 9.5 fue cercana a la solución verdadera, existe una pequeña discrepancia en el resultado de x_3 [ecuación (E9.5.10)]. Esta diferencia, que corresponde a un error relativo del –0.00043%, se debe al uso de seis cifras significativas durante los cálculos. Si se hubiesen utilizado más cifras significativas, el error en los resultados se habría reducido considerablemente. Si se hubiesen usado fracciones en lugar de decimales (y en consecuencia evitado los errores de redondeo), los resultados habrían sido exactos. Sin embargo, como las computadoras manejan sólo un número limitado de cifras significativas (recuerde la sección 3.4.1), es posible que ocurran errores de redondeo y se deben considerar al evaluar los resultados.

El problema de los errores de redondeo llega a volverse particularmente importante cuando se trata de resolver un gran número de ecuaciones. Esto se debe al hecho de que cada resultado depende del anterior. Por consiguiente, un error en los primeros pasos tiende a propagarse, es decir, a causar errores en los siguientes pasos.

Resulta complicado especificar el tamaño de los sistemas donde los errores de redondeo son significativos, ya que depende del tipo de computadora y de las propiedades de las ecuaciones. Una regla generalizada consiste en suponer que los errores de redondeo son de importancia cuando se trata de sistemas de 100 o más ecuaciones. En cualquier caso, siempre se deben sustituir los resultados en las ecuaciones originales y verificar si ha ocurrido un error sustancial. No obstante, como se verá más adelante, las magnitudes de los mismos coeficientes pueden influir en la aceptación de si una de estas pruebas de error asegura un resultado confiable.

9.3.3 Sistemas mal condicionados

Lo adecuado de una solución depende de la condición del sistema. En la sección 9.1.1 se desarrolló una representación gráfica de la condición de un sistema. Como se estudió en la sección 4.2.3, los

sistemas bien condicionados son aquellos en los que un pequeño cambio en uno o más coeficientes provoca un cambio similarmente pequeño en la solución. Los *sistemas mal condicionados* son aquellos en donde pequeños cambios en los coeficientes generan grandes cambios en la solución. Otra interpretación del mal condicionamiento es que un amplio rango de resultados puede satisfacer las ecuaciones en forma aproximada. Debido a que los errores de redondeo llegan a provocar pequeños cambios en los coeficientes, estos cambios artificiales pueden generar grandes errores en la solución de sistemas mal condicionados, como se ilustra en el siguiente ejemplo.

EJEMPLO 9.6 Sistemas mal condicionados

Planteamiento del problema Resuelva el siguiente sistema:

$$x_1 + 2x_2 = 10 \qquad \text{(E9.6.1)}$$

$$1.1x_1 + 2x_2 = 10.4 \qquad \text{(E9.6.2)}$$

Después, resuélvalo de nuevo, pero con el coeficiente x_1 de la segunda ecuación modificado ligeramente como 1.05.

Solución Usando las ecuaciones (9.10) y (9.11), la solución es

$$x_1 = \frac{2(10) - 2(10.4)}{1(2) - 2(1.1)} = 4$$

$$x_2 = \frac{1(10.4) - 1.1(10)}{1(2) - 2(1.1)} = 3$$

Sin embargo, con un ligero cambio al coeficiente a_{21} de 1.1 a 1.05, el resultado cambia de forma drástica a

$$x_1 = \frac{2(10) - 2(10.4)}{1(2) - 2(1.05)} = 8$$

$$x_2 = \frac{1(10.4) - 1.05(10)}{1(2) - 2(1.05)} = 1$$

Observe que la razón principal de la discrepancia entre los dos resultados es que el denominador representa la diferencia de dos números casi iguales. Como se explicó previamente en la sección 3.4.2, tales diferencias son altamente sensibles a pequeñas variaciones en los números empleados.

En este punto, se podría sugerir que la sustitución de los resultados en las ecuaciones originales alertaría al lector respecto al problema. Por desgracia, con frecuencia éste no es el caso en sistemas mal condicionados. La sustitución de los valores erróneos $x_1 = 8$ y $x_2 = 1$ en las ecuaciones (E9.6.1) y (E9.6.2) resulta en

$$8 + 2(1) = 10 = 10$$

$$1.1(8) + 2(1) = 10.8 \cong 10.4$$

Por lo tanto, aunque $x_1 = 8$ y $x_2 = 1$ no sea la solución verdadera al problema original, la prueba de error es lo suficientemente cercana para quizá confundirlo y hacerle creer que las soluciones son las adecuadas.

Como se hizo antes en la sección sobre métodos gráficos, es posible dar una representación visual del mal condicionamiento al graficar las ecuaciones (E9.6.1) y (E9.6.2) (recuerde la figura 9.2). Debido a que las pendientes de las líneas son casi iguales, visualmente es difícil percibir con exactitud dónde se intersecan. Dicha dificultad visual se refleja en forma cuantitativa en los resultados ambiguos del ejemplo 9.6. Esta situación se puede caracterizar matemáticamente escribiendo las dos ecuaciones en su forma general:

$$a_{11}x_1 + a_{12}x_2 = b_1 \qquad (9.24)$$

$$a_{21}x_1 + a_{22}x_2 = b_2 \tag{9.25}$$

Dividiendo la ecuación (9.24) entre a_{12} y la (9.25) entre a_{22}, y reordenando términos, se obtienen las versiones alternativas en el formato de líneas rectas [x_2 = (pendiente) x_1 + intercepto]:

$$x_2 = -\frac{a_{11}}{a_{12}}x_1 + \frac{b_1}{a_{12}}$$

$$x_2 = -\frac{a_{21}}{a_{22}}x_1 + \frac{b_2}{a_{22}}$$

Por consiguiente, si las pendientes son casi iguales

$$\frac{a_{11}}{a_{12}} \cong \frac{a_{21}}{a_{22}}$$

o, multiplicando en cruz,

$$a_{11}a_{22} \cong a_{12}a_{21}$$

lo cual se expresa también como

$$a_{11}a_{22} - a_{12}a_{21} \cong 0 \tag{9.26}$$

Ahora, si recordamos que $a_{11}a_{22} - a_{12}a_{21}$ es el determinante de un sistema bidimensional [ecuación (9.3)], se llega a la conclusión general de que un sistema mal condicionado es aquel en el que su determinante es cercano a cero. De hecho, si el determinante es exactamente igual a cero, las dos pendientes son idénticas, lo cual indica ya sea que no hay solución o que hay un número infinito de soluciones, como es el caso de los sistemas singulares ilustrados en las figuras 9.2*a* y 9.2*b*.

Es difícil especificar qué tan cerca de cero debe estar el determinante de manera que indique un mal condicionamiento. Esto se complica por el hecho de que el determinante puede cambiar al multiplicar una o más ecuaciones por un factor de escalamiento sin alterar la solución. Por consiguiente, el determinante es un valor relativo que se ve influenciado por la magnitud de los coeficientes.

EJEMPLO 9.7 Efecto de escalamiento sobre el determinante

Planteamiento del problema Evalúe el determinante de los siguientes sistemas:

a) Del ejemplo 9.1:

$$3x_1 + 2x_2 = 18 \tag{E9.7.1}$$
$$-x_1 + 2x_2 = 2 \tag{E9.7.2}$$

b) Del ejemplo 9.6:

$$x_1 + 2x_2 = 10 \tag{E9.7.3}$$
$$1.1x_1 + 2x_2 = 10.4 \tag{E9.7.4}$$

c) Repita *b*), pero multiplique las ecuaciones por 10.

Solución

a) El determinante de las ecuaciones (E9.7.1) y (E.9.7.2) que están bien condicionadas, es

$$D = 3(2) - 2(-1) = 8$$

b) El determinante de las ecuaciones (E9.7.3) y (E9.7.4), que están mal condicionadas, es

$$D = 1(2) - 2(1.1) = -0.2$$

c) Los resultados en *a*) *y b*) parecen corroborar el argumento de que los sistemas mal condicionados tienen determinantes cercanos a cero. Sin embargo, suponga que el sistema mal condicionado en *b*) se multiplica por 10, para obtener

$10x_1 + 20x_2 = 100$
$11x_1 + 20x_2 = 104$

La multiplicación de una ecuación por una constante no tiene efecto en su solución. Además, todavía está mal condicionada. Esto se verifica por el hecho de que multiplicar por una constante no tiene efecto en la solución gráfica. No obstante, el determinante se afecta en forma drástica:

$$D = 10(20) - 20(11) = -20$$

No sólo se han elevado en dos órdenes de magnitud, sino que ahora es más de dos veces el determinante del sistema bien condicionado en *a*).

Como se ilustró en el ejemplo anterior, la magnitud de los coeficientes interpone un efecto de escalamiento, que complica la relación entre la condición del sistema y el tamaño del determinante. Una manera de evitar parcialmente esta dificultad es escalando las ecuaciones en forma tal que el elemento máximo en cualquier renglón sea igual a 1.

EJEMPLO 9.8 Escalamiento

Planteamiento del problema Escale los sistemas de ecuaciones del ejemplo 9.7 a un valor máximo de 1 y calcule de nuevo sus determinantes.

Solución

a) Para el sistema bien condicionado, el escalamiento resulta en

$$x_1 + 0.667x_2 = 6$$
$$-0.5x_1 + x_2 = 1$$

cuyo determinante es

$$D = 1(1) - 0.667(-0.5) = 1.333$$

b) Para el sistema mal condicionado, el escalamiento resulta en

$$0.5x_1 + x_2 = 5$$
$$0.55x_1 + x_2 = 5.2$$

cuyo determinante es

$$D = 0.5(1) - 1(0.55) = -0.05$$

c) En el último caso, al realizar los cambios del escalamiento, el sistema toma la misma forma que en *b*) y el determinante es también –0.05. De esta forma, se remueve el efecto de escalamiento.

En una sección anterior (sección 9.1.2) se mencionó que el determinante es difícil de evaluar para más de tres ecuaciones simultáneas. Por lo tanto, podría parecer que no ofrece un recurso práctico para evaluar la condición de un sistema. Sin embargo, como se describe en el cuadro 9.1, existe un algoritmo simple que resulta de la eliminación de Gauss y que se puede usar para la evaluación del determinante.

Además del método usado en el ejemplo anterior existen otras formas para evaluar la condición del sistema. Por ejemplo, hay métodos alternativos para normalizar los elementos (véase Stark, 1970). Además, como se verá en el capítulo siguiente (sección 10.3), la matriz inversa y la norma de una matriz pueden usarse para evaluar la condición de un sistema. Por último, una prueba simple (pero que consume tiempo) consiste en modificar ligeramente los coeficientes y repetir la solución. Si tales modificaciones generan resultados drásticamente diferentes, es posible que el sistema esté mal condicionado.

Como se deduce del análisis anterior, los sistemas mal condicionados resultan problemáticos. Por fortuna, la mayoría de las ecuaciones algebraicas lineales, obtenidas de un problema de ingeniería, son por naturaleza bien condicionadas. Además, algunas de las técnicas presentadas en la sección 9.4 ayudarán a reducir el problema.

9.3.4 Sistemas singulares

En la sección anterior se aprendió que una forma con la cual un sistema de ecuaciones puede estar mal condicionado es cuando dos o más de las ecuaciones son casi idénticas. Obviamente aún es peor cuando las dos son idénticas. En tales casos, se pierde un grado de libertad y se daría un caso imposible de $n - 1$ ecuaciones con n incógnitas. Tales casos podrían no ser obvios, en particular cuando se enfrenta con grandes sistemas de ecuaciones. En consecuencia, sería útil tener una forma de detectar la singularidad de manera automática.

La respuesta a este problema está claramente dada por el hecho de que el determinante de un sistema singular es cero. Esta idea, a su vez, puede relacionarse con la eliminación gaussiana reconociendo que después del paso de eliminación, el determinante se evalúa como el producto de los elementos de la diagonal (recuerde el cuadro 9.1). Así, un algoritmo de computadora puede efectuar una prueba para discernir si se crea un cero en la diagonal durante la etapa de la eliminación. Si se descubre uno, el cálculo se puede terminar inmediatamente y en la pantalla aparecerá un mensaje de alerta. Se mostrarán más tarde, en este capítulo, los detalles de cómo se realiza esto cuando se presente el algoritmo completo de la eliminación de Gauss.

9.4 TÉCNICAS PARA MEJORAR LAS SOLUCIONES

Las siguientes técnicas se pueden incorporar al algoritmo de eliminación de Gauss simple, para evitar algunos de los problemas analizados en la sección previa.

Cuadro 9.1 Evaluación de determinantes usando la eliminación de Gauss

En la sección 9.1.2 se dijo que la evaluación de los determinantes por expansión de menores no resultaba práctico para grandes sistemas de ecuaciones. De esta forma, se concluyó que la regla de Cramer sólo es aplicable a sistemas pequeños. Sin embargo, como se mencionó en la sección 9.3.3, el valor del determinante permite estimar la condición de un sistema. Por lo tanto, será útil tener un método práctico para calcular esta cantidad.

Por fortuna, la eliminación gaussiana proporciona una forma simple para hacerlo. El método se basa en el hecho de que el determinante de una matriz triangular se puede calcular de forma simple, como el producto de los elementos de su diagonal:

$$D = a_{11}a_{22}a_{33} \ldots a_{nn} \tag{C9.1.1}$$

La validez de esta formulación se ilustra para un sistema de 3 por 3:

$$D = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{vmatrix}$$

donde el determinante se evalúa como [recuerde la ecuación (9.4)]

$$D = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ 0 & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} 0 & a_{23} \\ 0 & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} 0 & a_{22} \\ 0 & 0 \end{vmatrix}$$

o evaluando los menores (es decir, los determinantes 2 por 2)

$$D = a_{11}a_{22}a_{23} - a_{12}(0) + a_{13}(0) = a_{11}a_{22}a_{33}$$

Recuerde que el paso de eliminación hacia adelante de la eliminación de Gauss genera un sistema triangular superior. Puesto que el valor del determinante no cambia con el proceso de eliminación hacia adelante, simplemente el determinante se evalúa al final de este paso por medio de

$$D = a_{11}a'_{22}a''_{33} \cdots a_{nn}^{(n-1)} \tag{C9.1.2}$$

donde los superíndices indican el número de veces que los elementos han sido modificados en el proceso de eliminación. Por lo tanto, es posible aprovechar el esfuerzo que se ha logrado al reducir el sistema a su forma triangular, y obtener un cálculo simple del determinante.

Hay una ligera modificación al método anterior cuando el programa usa pivoteo parcial (la sección 9.4.2). En este caso, el determinante cambia de signo cada vez que un renglón es pivoteado. Una manera de representar esto es modificando la ecuación (C9.1.2):

$$D = a_{11}a'_{22}a''_{33} \cdots a_{nn}^{(n-1)}(-1)^p \tag{C9.1.3}$$

donde p representa el número de veces en que los renglones se pivotean. Esta modificación se puede incorporar de forma simple en un programa; únicamente rastree el número de pivoteos que se llevan a cabo durante el transcurso de los cálculos y después use la ecuación (C9.1.3) para evaluar el determinante.

9.4.1 Uso de más cifras significativas

El remedio más simple para el mal condicionamiento consiste en emplear más cifras significativas en los cálculos. Si la computadora tiene la capacidad para usar más cifras, esta característica reducirá enormemente el problema. No obstante, el precio que hay que pagar en cálculo y memoria se eleva con el uso de la precisión extendida (recuerde la sección 3.4.1).

9.4.2 Pivoteo

Como se mencionó al inicio de la sección 9.3, ocurren problemas obvios cuando un elemento pivote es cero, ya que el paso de normalización origina una división entre cero. También llegan a surgir problemas cuando el elemento pivote es cercano a —o más aún que sea exactamente igual a— cero, debido a que si la magnitud del elemento pivote es pequeña comparada con los otros elementos, entonces se pueden introducir errores de redondeo.

Por lo tanto, antes de normalizar cada renglón, resulta conveniente determinar el coeficiente más grande disponible en la columna debajo del elemento pivote. Los renglones se pueden intercambiar de manera que el elemento más grande sea el elemento pivote; esto se conoce como *pivoteo parcial*. Al procedimiento, donde tanto en las columnas como en los renglones se busca el elemento más grande y luego se intercambian, se le conoce como *pivoteo completo*, el cual se usa en muy raras ocasiones debido a que al intercambiar columnas se cambia el orden de las x y, en consecuencia, se agrega complejidad significativa y usualmente injustificada al programa de computadora. El siguiente ejemplo ilustra las ventajas del pivoteo parcial. Además de evitar la división entre cero, el pivoteo también minimiza el error de redondeo. Como tal, sirve también para resolver parcialmente el mal condicionamiento.

EJEMPLO 9.9 Pivoteo parcial

Planteamiento del problema Emplee la eliminación de Gauss para resolver

$$0.0003x_1 + 3.0000x_2 = 2.0001$$
$$1.0000x_1 + 1.0000x_2 = 1.0000$$

Observe que en esta forma el primer elemento pivote, $a_{11} = 0.0003$, es muy cercano a cero. Entonces haga de nuevo el cálculo, pero ahora con pivoteo parcial, invirtiendo el orden de las ecuaciones. La solución exacta es $x_1 = 1/3$ y $x_2 = 2/3$.

Solución Multiplicando la primera ecuación por 1/(0.0003) da como resultado

$$x_1 + 10\,000x_2 = 6\,667$$

lo cual se utiliza para eliminar x_1 de la segunda ecuación:

$$-9\,999x_2 = -6\,666$$

de donde se despeja

$$x_2 = \frac{2}{3}$$

Este resultado se sustituye en la primera ecuación para evaluar x_1:

$$x_1 = \frac{2.0001 - 3(2/3)}{0.0003} \tag{E9.9.1}$$

Sin embargo, debido a la cancelación por resta, el resultado es muy sensible al número de cifras significativas empleadas en el cálculo:

Cifras significativas	x_2	x_1	Valor absoluto del error relativo porcentual para x_1
3	0.667	–3.33	1 099
4	0.6667	0.0000	100
5	0.66667	0.30000	10
6	0.666667	0.330000	1
7	0.6666667	0.3330000	0.1

Observe cómo el valor de x_1 depende en gran medida del número de cifras significativas. Esto se debe a que en la ecuación (E9.9.1) se restan dos números casi iguales. Por otro lado, si se resuelven las ecuaciones en orden inverso, se normaliza el renglón con el elemento pivote más grande. Las ecuaciones son

$$1.0000x_1 + 1.0000x_2 = 1.0000$$
$$0.0003x_1 + 3.0000x_2 = 2.0001$$

La eliminación y la sustitución dan $x_2 = 2/3$. Con diferentes números de cifras significativas, x_1 se puede calcular de la primera ecuación, así

$$x_1 = \frac{1-(2/3)}{1} \tag{E9.9.2}$$

Este caso es mucho menos sensible al número de cifras significativas usadas en el cálculo:

Cifras significativas	x_2	x_1	Valor absoluto del error relativo porcentual para x_1
3	0.667	0.333	0.1
4	0.6667	0.3333	0.01
5	0.66667	0.33333	0.001
6	0.666667	0.333333	0.0001
7	0.6666667	0.3333333	0.00001

Por lo que la estrategia de pivoteo es mucho más satisfactoria.

```
p = k
big = |a_k,k|
DOFOR ii = k+1, n
  dummy = |a_ii,k|
  IF (dummy > big)
     big = dummy
     p = ii
  END IF
END DO
IF (p ≠ k)
  DOFOR jj = k, n
     dummy = a_p,jj
     a_p,jj = a_k,jj
     a_k,jj = dummy
  END DO
  dummy = b_p
  b_p = b_k
  b_k = dummy
END IF
```

Figura 9.5
Pseudocódigo para implementar el pivoteo parcial.

Los programas computacionales de uso general deben tener una estrategia de pivoteo. En la figura 9.5 se proporciona un algoritmo simple para llevar a cabo dicha estrategia. Observe que el algoritmo consiste en dos grandes ciclos. Luego de guardar el elemento pivote actual y su número de renglón como las variables *big* y *p*, el primer ciclo compara el elemento pivote con los elementos que se hallan debajo de él, para verificar si algunos de ellos es mayor que el elemento pivote. Si es así, el nuevo elemento más grande y su número de renglón se guardan en *big* y *p*. Después, el segundo ciclo intercambia el renglón del pivote original con el del elemento más grande, de tal forma que el último sea el nuevo renglón pivote. Este pseudocódigo puede agregarse a un programa basado en los otros elementos de la eliminación de Gauss mostrados en la figura 9.4. La mejor forma de hacerlo consiste en emplear un método modular y escribir la figura 9.5 como una subrutina (o procedimiento), que pueda llamarse directamente después del inicio del primer ciclo en la figura 9.4*a*.

Observe que la segunda instrucción IF/THEN de la figura 9.5 intercambia físicamente los renglones. Con grandes matrices, esto llevaría mucho tiempo. En consecuencia, de hecho, la mayoría de los códigos no intercambian renglones sino llevan un registro de cuál es el renglón pivote, guardando los subíndices apropiados en un vector. Este vector proporciona luego una base para especificar el orden adecuado de los renglones durante la eliminación hacia adelante y las operaciones de sustitución hacia atrás. Así, se dice que las operaciones se implementan *in situ*.

9.4.3 Escalamiento

En la sección 9.3.3 se mencionó que el escalamiento podía ser útil para la estandarización del tamaño determinante. Más allá de esta aplicación, tiene utilidad en la minimización de los errores de

redondeo, en aquellos casos en los que algunas de las ecuaciones de un sistema tienen coeficientes mucho más grandes que otros. Tales situaciones se encuentran con frecuencia en la práctica de la ingeniería, al usar unidades muy diferentes en el desarrollo de ecuaciones simultáneas. Por ejemplo, en problemas de circuitos eléctricos, los voltajes desconocidos se pueden expresar en unidades que varían desde microvoltios hasta kilovoltios. Existen ejemplos similares en todos los campos de la ingeniería. Mientras cada una de las ecuaciones sea consistente, el sistema será técnicamente correcto y susceptible de ser resuelto. Sin embargo, el uso de unidades tan diversas puede llevar a que los coeficientes difieran ampliamente en magnitud. Esto, a su vez, puede tener un impacto sobre el error de redondeo, ya que afecta el pivoteo, como se ilustra en el siguiente ejemplo.

EJEMPLO 9.10 Efecto del escalamiento sobre el pivoteo y el redondeo

Planteamiento del problema

a) Resuelva el siguiente sistema de ecuaciones usando la eliminación de Gauss y una estrategia de pivoteo:

$$2x_1 + 100\,000x_2 = 100\,000$$
$$x_1 + x_2 = 2$$

b) Repita el problema después de escalar las ecuaciones de tal forma que el coeficiente máximo en cada renglón sea 1.

c) Finalmente, utilice los coeficientes escalados para determinar si el pivoteo es necesario. No obstante, resuelva las ecuaciones con los valores de los coeficientes originales. En todos los casos, conserve sólo tres cifras significativas. Observe que las respuestas correctas son x_1 = 1.00002 y x_2 = 0.99998 o, para tres cifras significativas, $x_1 = x_2 = 1.00$.

Solución

a) Sin escalar, se aplica la eliminación hacia adelante y se obtiene

$$2x_1 + 100\,000x_2 = 100\,000$$
$$-50\,000x_2 = -50\,000$$

que se puede resolver por sustitución hacia atrás:

$$x_2 = 1.00$$
$$x_1 = 0.00$$

Aunque x_2 es correcta, x_1 tiene un 100% de error debido al redondeo.

b) El escalamiento transforma las ecuaciones originales en

$$0.00002x_1 + x_2 = 1$$
$$x_1 + x_2 = 2$$

Por lo tanto, se deben pivotear los renglones y colocar el valor más grande sobre la diagonal.

$$x_1 + x_2 = 2$$
$$0.00002x_1 + x_2 = 1$$

La eliminación hacia adelante da como resultado

$$x_1 + x_2 = 2$$
$$x_2 = 1.00$$

de donde se obtiene

$$x_1 = x_2 = 1$$

De esta forma, el escalamiento conduce a la respuesta correcta.

c) Los coeficientes escalados revelan que es necesario el pivoteo. Por lo tanto, se pivotea pero se mantienen los coeficientes originales para obtener

$$x_1 + \quad x_2 = 2$$
$$2x_1 + 100\,000x_2 = 100\,000$$

La eliminación hacia adelante da como resultado

$$x_1 + \quad x_2 = 2$$
$$100\,000x_2 = 100\,000$$

que al resolverse se obtiene la respuesta correcta: $x_1 = x_2 = 1$. Entonces, el escalamiento fue útil para determinar si el pivoteo era necesario; aunque las ecuaciones por sí mismas no requieren escalarse para llegar a un resultado correcto.

Al igual que en el ejemplo anterior, el escalamiento es útil para minimizar los errores de redondeo. Sin embargo, se debe advertir que el propio escalamiento lleva también a errores de redondeo. Por ejemplo, dada la ecuación

$$2x_1 + 300\,000x_2 = 1$$

y usando tres cifras significativas, escalando se obtiene

$$0.00000667x_1 + x_2 = 0.00000333$$

De esta forma, el escalamiento introduce un error de redondeo en el primer coeficiente y en la constante del lado derecho. Por esta razón, algunas veces se sugiere que el escalamiento se emplee únicamente como en el inciso *c*) del ejemplo anterior. Esto es, se usa para calcular valores escalados de los coeficientes sólo como un criterio de pivoteo; pero los valores de los coeficientes originales se conservan para los cálculos reales de eliminación y sustitución. Esto tiene ventajas y desventajas si el determinante se calcula como parte del programa. Es decir, el determinante resultante no será escalado. Sin embargo, como muchas aplicaciones de la eliminación de Gauss no requieren la evaluación del determinante, es el planteamiento más común y se usará en el algoritmo de la siguiente sección.

9.4.4 Algoritmo para la eliminación gaussiana

Los algoritmos de las figuras 9.4 y 9.5 se combinan ahora en un solo algoritmo para implementar el algoritmo completo de la eliminación de Gauss. En la figura 9.6 se muestra el algoritmo de una subrutina general para realizar la eliminación de Gauss.

Observe que el programa tiene módulos para las tres operaciones principales del algoritmo de eliminación gaussiana: eliminación hacia adelante, sustitución hacia atrás y pivoteo. Además, hay varios aspectos del código que difieren y representan un mejoramiento de los pseudocódigos de las figuras 9.4 y 9.5. Éstos son:

- Las ecuaciones no están escaladas, pero los valores escalados de los elementos se usan para determinar si se debe usar el pivoteo.
- El término diagonal se vigila durante la fase del pivoteo para detectar ocurrencias de valores cercanos a cero y con esto indicar si el sistema es singular. Si devuelve un valor de $er = -1$, se ha detectado una matriz singular y el cálculo debe terminar. El usuario da a un parámetro *tol*, un número pequeño para detectar ocurrencias cercanas a cero.

EJEMPLO 9.11 Solución de ecuaciones algebraicas lineales por medio de la computadora

Planteamiento del problema Un programa de computadora para resolver ecuaciones algebraicas lineales, como por ejemplo en el que se basa la figura 9.6, sirve para resolver un problema relacionado con el ejemplo de la caída del paracaidista, analizado en el capítulo 1. Suponga que un equipo de tres paracaidistas está unido por una cuerda ligera mientras va en caída libre a una velocidad de 5 m/s (figura 9.7). Calcule la tensión en cada sección de la cuerda y la aceleración del equipo, dados los siguientes datos:

Paracaidista	Masa, kg	Coeficiente de arrastre, kg/s
1	70	10
2	60	14
3	40	17

Solución Los diagramas de cuerpo libre para cada paracaidista se muestran en la figura 9.8. Sumando las fuerzas en la dirección vertical y utilizando la segunda ley de Newton se obtiene un sistema de tres ecuaciones lineales simultáneas:

$$m_1g - T - c_1v \qquad = m_1a$$
$$m_2g + T - c_2v - R = m_2a$$
$$m_3g \qquad - c_3v + R = m_3a$$

Estas ecuaciones tienen tres incógnitas: a, T y R. Después de sustituir los valores conocidos, las ecuaciones se pueden expresar en forma matricial como (g = 9.8 m/s²),

```
SUB Gauss (a, b, n, x, tol, er)
  DIMENSION s (n)
  er = 0
  DOFOR i = 1, n
    s_i = ABS(a_i,1)
    DOFOR j = 2, n
      IF ABS(a_i,j)>s_i THEN s_i = ABS(a_i,j)
    END DO
  END DO
  CALL Eliminate(a, s, n, b, tol, er)
  IF er ≠ −1 THEN
     CALL Substitute(a, n, b, x)
  END IF
END Gauss

SUB Eliminate (a, s, n, b, tol, er)
  DOFOR k = 1, n − 1
    CALL Pivot (a, b, s, n, k)
    IF ABS (a_k,k/s_k) < tol THEN
      er = −1
      EXIT DO
    END IF
    DOFOR i = k + 1, n
      factor = a_i,k/a_k,k
      DOFOR j = k + 1, n
        a_i,j = a_i,j − factor*a_k,j
      END DO
      b_i = b_i - factor * b_k
    END DO
  END DO
  IF ABS(a_n,n/s_n) < to1 THEN er = −1
END Eliminate
```

```
SUB Pivot (a, b, s, n, k)
  p = k
  big = ABS(a_k,k/s_k)
  DOFOR ii = k + 1, n
    dummy = ABS(a_ii,k/s_ii)
    IF dummy > big THEN
      big = dummy
      p = ii
    END IF
  END DO
  IF p ≠ k THEN
     DOFOR jj = k, n
      dummy = a_p,jj
      a_p,jj = a_k,jj
      a_k,jj = dummy
     END DO
     dummy = b_p
     b_p = b_k
     b_k = dummy
     dummy = s_p
     s_p = s_k
     s_k = dummy
  END IF
END pivot

SUB Substitute (a, n, b, x)
  x_n = b_n/a_n,n
  DOFOR i = n − 1, 1, −1
    sum = 0
    DOFOR j = i + 1, n
     sum = sum + a_i,j * x_j
    END DO
    x_n = (b_n − sum) / a_n,n
  END DO
END Substitute
```

Figura 9.6
Pseudocódigo para instaurar la eliminación de Gauss con pivoteo parcial.

$$\begin{bmatrix} 70 & 1 & 0 \\ 60 & -1 & 1 \\ 40 & 0 & -1 \end{bmatrix} \begin{Bmatrix} a \\ T \\ R \end{Bmatrix} = \begin{Bmatrix} 636.7 \\ 518.6 \\ 307.4 \end{Bmatrix}$$

Este sistema se puede resolver usando sus propios programas. El resultado es $a = 8.6041$ m/s^2, $T = 34.4118$ N y $R = 36.7647$ N.

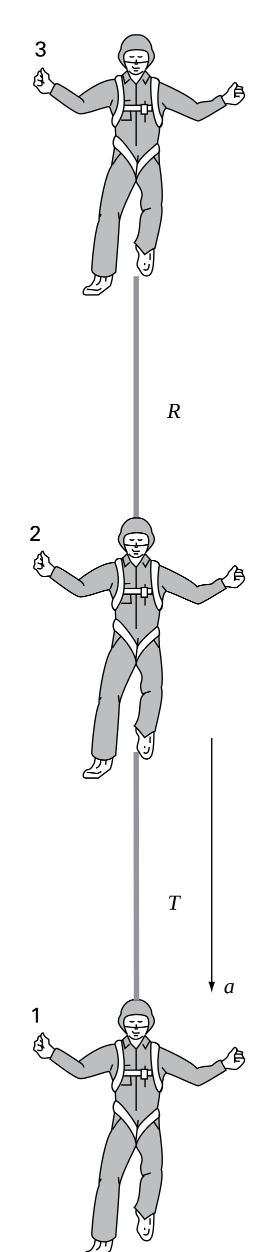


Figura 9.7
Tres paracaidistas en caída libre unidos por cuerdas sin peso.

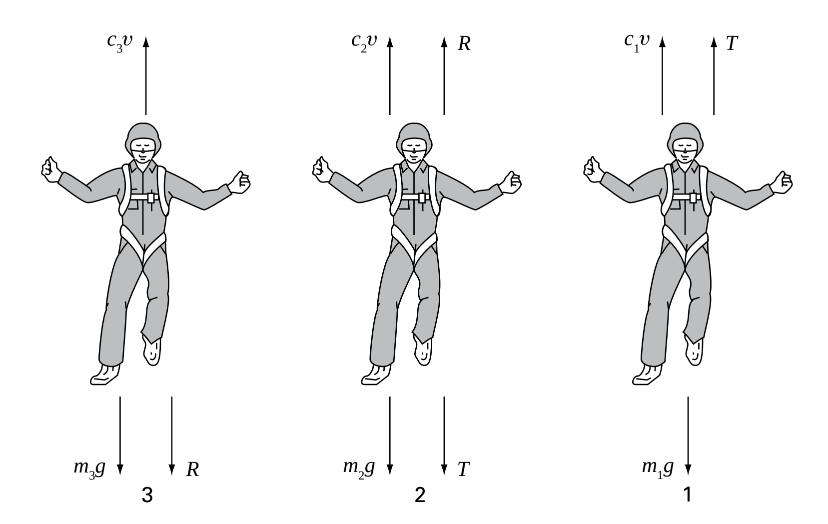


Figura 9.8
Diagramas de cuerpo libre para cada uno de los tres paracaidistas en caída.

9.5 SISTEMAS COMPLEJOS

En algunos problemas es posible obtener un sistema de ecuaciones complejas

$$[C]\{Z\} = \{W\} \tag{9.27}$$

donde

$$[C] = [A] + i[B]$$

$$\{Z\} = \{X\} + i\{Y\}$$

$$\{W\} = \{U\} + i\{V\} \tag{9.28}$$

donde $i = \sqrt{-1}$.

El camino más directo para resolver un sistema como éste consiste en emplear uno de los algoritmos descritos en esta parte del libro; pero sustituyendo todas las operaciones reales por complejas. Claro que esto sólo es posible con aquellos lenguajes, como el Fortran, que permiten el uso de variables complejas.

Para lenguajes que no permiten la declaración de variables complejas, es posible escribir un código que convierta operaciones reales en complejas. Sin embargo, esto no es una tarea trivial. Una alternativa es convertir el sistema complejo en uno equivalente que trabaje con variables reales. Esto se logra al sustituir la ecuación (9.28) en la (9.27) e igualar las partes real y compleja de la ecuación resultante, para obtener

$$[A]\{X\} - [B]\{Y\} = \{U\} \tag{9.29}$$

y

$$[B]\{X\} + [A]\{Y\} = \{V\} \tag{9.30}$$

Así, el sistema de n ecuaciones complejas se convierte en un conjunto de $2n$ ecuaciones reales. Esto significa que el tiempo de almacenamiento y de ejecución se incrementará en forma significativa. En consecuencia, habrá que evaluar las ventajas y desventajas de esta opción. Si es poco frecuente que se evalúen sistemas complejos, es preferible usar las ecuaciones (9.29) y (9.30) por su conveniencia. Sin embargo, si se usan con frecuencia y desea utilizar un lenguaje que no permite el uso de datos de tipo complejo, quizá valga la pena escribir un programa que convierta operaciones reales en complejas.

9.6 SISTEMAS DE ECUACIONES NO LINEALES

Recuerde que al final del capítulo 6 se expuso un procedimiento para resolver dos ecuaciones no lineales con dos incógnitas. Éste se puede extender al caso general para resolver n ecuaciones no lineales simultáneas.

$$\begin{aligned} f_1(x_1, x_2, \ldots, x_n) &= 0 \\ f_2(x_1, x_2, \ldots, x_n) &= 0 \\ &\vdots \\ f_n(x_1, x_2, \ldots, x_n) &= 0 \end{aligned} \tag{9.31}$$

La solución de este sistema consiste en un conjunto de valores x que hacen todas las ecuaciones igual a cero.

Como se describió en la sección 6.5.2, un procedimiento para resolver tales sistemas se basa en la versión multidimensional del método de Newton-Raphson. Así, se escribe para cada ecuación una expansión de la serie de Taylor. Por ejemplo, para la k-ésima ecuación,

$$f_{k,i+1} = f_{k,i} + (x_{1,i+1} - x_{1,i})\frac{\partial f_{k,i}}{\partial x_1} + (x_{2,i+1} - x_{2,i})\frac{\partial f_{k,i}}{\partial x_2} + \cdots + (x_{n,i+1} - x_{n,i})\frac{\partial f_{k,i}}{\partial x_n} \tag{9.32}$$

donde el primer subíndice, k, representa la ecuación o la incógnita, y el segundo subíndice denota si el valor de la función en cuestión es el presente (i) o el siguiente ($i + 1$).

Las ecuaciones de la forma (9.32) son escritas para cada una de las ecuaciones no lineales originales. Después, como se hizo al obtener la ecuación (6.20) a partir de la (6.19), todos los términos $f_{k,i+1}$ se igualan a cero, como sería el caso en la raíz, y la ecuación (9.32) se escribe como

$$\begin{aligned} &-f_{k,i} + x_{1,i}\frac{\partial f_{k,i}}{\partial x_1} + x_{2,i}\frac{\partial f_{k,i}}{\partial x_2} + \cdots + x_{n,i}\frac{\partial f_{k,i}}{\partial x_n} \\ &= x_{1,i+1}\frac{\partial f_{k,i}}{\partial x_1} + x_{2,i+1}\frac{\partial f_{k,i}}{\partial x_2} + \cdots + x_{n,i+1}\frac{\partial f_{k,i}}{\partial x_n} \end{aligned} \tag{9.33}$$

Observe que las únicas incógnitas en la ecuación (9.33) son los términos $x_{k,i+1}$ del lado derecho. Todas las otras cantidades tienen su valor presente (i) y, por lo tanto, son conocidas en cualquier iteración. En consecuencia, el sistema de ecuaciones representado, en general, por la ecuación (9.33) (es decir, con $k = 1, 2, \ldots, n$) constituye un sistema de ecuaciones lineales simultáneas que se pueden resolver con los métodos analizados en esta parte del libro.

Se puede emplear la notación matricial para expresar la ecuación (9.33) en forma concisa. Las derivadas parciales se expresan como

$$[Z] = \begin{bmatrix} \frac{\partial f_{1,i}}{\partial x_1} & \frac{\partial f_{1,i}}{\partial x_2} & \cdots & \frac{\partial f_{1,i}}{\partial x_n} \\ \frac{\partial f_{2,i}}{\partial x_1} & \frac{\partial f_{2,i}}{\partial x_2} & \cdots & \frac{\partial f_{2,i}}{\partial x_n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \frac{\partial f_{n,i}}{\partial x_1} & \frac{\partial f_{n,i}}{\partial x_2} & \cdots & \frac{\partial f_{n,i}}{\partial x_n} \end{bmatrix} \tag{9.34}$$

Los valores inicial y final se expresan en forma vectorial como

$$\{X_i\}^T = \lfloor x_{1,i} \quad x_{2,i} \quad \ldots \quad x_{n,i} \rfloor$$

y

$$\{X_{i+1}\}^T = \lfloor x_{1,i+1} \quad x_{2,i+1} \quad \ldots \quad x_{n,i+1} \rfloor$$

Finalmente, los valores de la función en i se pueden expresar como

$$\{F_i\}^T = \lfloor f_{1,i} \quad f_{2,i} \quad \cdots \quad f_{n,i} \rfloor$$

Usando estas relaciones, la ecuación (9.33) se representa en forma concisa como

$$[Z]\{X_{i+1}\} = -\{F_i\} + [Z]\{X_i\} \tag{9.35}$$

La ecuación (9.35) se resuelve usando una técnica como la eliminación de Gauss. Este proceso se repite iterativamente para obtener una aproximación refinada de forma similar al caso de dos ecuaciones como en la sección 6.5.2.

Se debe notar que el procedimiento anterior tiene dos desventajas importantes. Primero, a menudo no es fácil evaluar la ecuación (9.34). Por lo que se ha desarrollado una variación del método de Newton-Raphson para evitar tal problema. Como podría esperarse, tal variación se basa en el uso de aproximaciones por diferencias finitas, para calcular las derivadas parciales que aparecen en $[Z]$.

La segunda desventaja del método de Newton-Raphson para multiecuaciones es que usualmente se requiere de excelentes valores iniciales para asegurar la convergencia. Ya que con frecuencia esto es difícil de obtener, se han desarrollado métodos alternos que, aunque son más lentos que el método de Newton-Raphson, dan un mejor comportamiento de convergencia. Un método común es reformular el sistema no lineal como una sola función

$$F(x) = \sum_{i=1}^{n} [f_i(x_1, x_2, \ldots, x_n)]^2 \tag{9.36}$$

donde $f_i(x_1, x_2, \ldots, x_n)$ es el i-ésimo miembro del sistema original de la ecuación (9.31). Los valores de x que minimizan esta función representan también la solución del sistema no lineal. Como se verá en el capítulo 17, esta reformulación pertenece a una clase de problemas llamados *regresión no lineal*. Como tal, se puede abordar con varias técnicas de optimización como las que se describirán más adelante en este texto (parte cuatro, específicamente en el capítulo 14).

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \vdots & b_1 \\ a_{21} & a_{22} & a_{23} & \vdots & b_2 \\ a_{31} & a_{32} & a_{33} & \vdots & b_3 \end{bmatrix}$$

$$\downarrow$$

$$\begin{bmatrix} 1 & 0 & 0 & \vdots & b_1^{(n)} \\ 0 & 1 & 0 & \vdots & b_2^{(n)} \\ 0 & 0 & 1 & \vdots & b_3^{(n)} \end{bmatrix}$$

$$\downarrow$$

$$\begin{aligned} x_1 \qquad\qquad &= b_1^{(n)} \\ x_2 \qquad &= b_2^{(n)} \\ x_3 &= b_3^{(n)} \end{aligned}$$

Figura 9.9
Ilustración del método Gauss-Jordan. Compare con la figura 9.3 para dilucidar las diferencias entre esta técnica y la eliminación de Gauss. El superíndice n significa que los elementos del vector del lado derecho se han modificado n veces (para este caso, $n = 3$).

9.7 GAUSS-JORDAN

El método de Gauss-Jordan es una variación de la eliminación de Gauss. La principal diferencia consiste en que cuando una incógnita se elimina en el método de Gauss-Jordan, ésta es eliminada de todas las otras ecuaciones, no sólo de las subsecuentes. Además, todos los renglones se normalizan al dividirlos entre su elemento pivote. De esta forma, el paso de eliminación genera una matriz identidad en vez de una triangular (figura 9.9). En consecuencia, no es necesario usar la sustitución hacia atrás para obtener la solución. El método se ilustra mejor con un ejemplo.

EJEMPLO 9.12 Método de Gauss-Jordan

Planteamiento del problema Con la técnica de Gauss-Jordan resuelva el sistema del ejemplo 9.5:

$$3x_1 - 0.1x_2 - 0.2x_3 = 7.85$$
$$0.1x_1 + 7x_2 - 0.3x_3 = -19.3$$
$$0.3x_1 - 0.2x_2 + 10x_3 = 71.4$$

Solución Primero, exprese los coeficientes y el lado derecho como una matriz aumentada:

$$\begin{bmatrix} 3 & -0.1 & -0.2 & 7.85 \\ 0.1 & 7 & -0.3 & -19.3 \\ 0.3 & -0.2 & 10 & 71.4 \end{bmatrix}$$

Luego normalice el primer renglón, dividiéndolo entre el elemento pivote, 3, para obtener

$$\begin{bmatrix} 1 & -0.0333333 & -0.066667 & 2.61667 \\ 0.1 & 7 & -0.3 & -19.3 \\ 0.3 & -0.2 & 10 & 71.4 \end{bmatrix}$$

El término x_1 se elimina del segundo renglón restando 0.1 veces al primer renglón del segundo. En forma similar, restando 0.3 veces el primer renglón del tercero, se eliminará el término x_1 del tercer renglón:

$$\begin{bmatrix} 1 & -0.0333333 & -0.066667 & 2.61667 \\ 0 & 7.00333 & -0.293333 & -19.5617 \\ 0 & -0.190000 & 10.0200 & 70.6150 \end{bmatrix}$$

En seguida, se normaliza el segundo renglón dividiéndolo entre 7.00333:

$$\begin{bmatrix} 1 & -0.0333333 & -0.066667 & 2.61667 \\ 0 & 1 & -0.0418848 & -2.79320 \\ 0 & -0.190000 & 10.0200 & 70.6150 \end{bmatrix}$$

Al reducir los términos x_2 de las ecuaciones primera y tercera se obtiene

$$\begin{bmatrix} 1 & 0 & -0.0680629 & 2.52356 \\ 0 & 1 & -0.0418848 & -2.79320 \\ 0 & 0 & 10.01200 & 70.0843 \end{bmatrix}$$

El tercer renglón se normaliza después al dividirlo entre 10.0120:

$$\begin{bmatrix} 1 & 0 & -0.0680629 & 2.52356 \\ 0 & 1 & -0.0418848 & -2.79320 \\ 0 & 0 & 1 & 7.0000 \end{bmatrix}$$

Por último, los términos x_3 se pueden eliminar de la primera y segunda ecuación para obtener

$$\begin{bmatrix} 1 & 0 & 0 & 3.0000 \\ 0 & 1 & 0 & -2.5000 \\ 0 & 0 & 1 & 7.0000 \end{bmatrix}$$

De esta forma, como se muestra en la figura 9.9, la matriz de coeficientes se ha transformado en la matriz identidad, y la solución se obtiene en el vector del lado derecho. Observe que no se requiere la sustitución hacia atrás para llegar a la solución.

Todo el material en este capítulo respecto a los escollos y las mejoras en la eliminación de Gauss también se aplica al método Gauss-Jordan. Por ejemplo, se puede usar una estrategia de pivoteo similar para evitar la división entre cero y para reducir el error de redondeo.

Aunque la técnica de Gauss-Jordan y la eliminación de Gauss podrían parecer casi idénticas, la primera requiere más trabajo. Con el empleo de un enfoque similar al de la sección 9.2.1, se determina que el número de flops que se involucra en la técnica de Gauss-Jordan simple es

$$n^3 + n^2 - n \xrightarrow[\text{aumenta}]{\text{conforme } n} n^3 + O(n^2) \tag{9.37}$$

Así, la técnica de Gauss-Jordan involucra aproximadamente 50% más operaciones que la eliminación de Gauss [compárese con la ecuación (9.23)]. Por lo tanto, la eliminación de Gauss es el método de eliminación sencilla que se prefiere para obtener las soluciones de ecuaciones algebraicas lineales. Sin embargo, una de las razones principales por las que se ha introducido la técnica de Gauss-Jordan, es que aún se utiliza tanto en la ingeniería como en ciertos algoritmos numéricos.

9.8 RESUMEN

En resumen, se ha dedicado la mayor parte de este capítulo a la eliminación de Gauss: el método fundamental para resolver ecuaciones algebraicas lineales simultáneas. Aunque es una de las técnicas más antiguas concebidas para este propósito, sin embargo, es un algoritmo efectivo en extremo para obtener las soluciones de muchos problemas en ingeniería. Además de esta utilidad práctica, este capítulo proporciona un contexto para el análisis de puntos generales, como el redondeo, el escalamiento y el condicionamiento. Se presentó también, en forma breve, material sobre el método de Gauss-Jordan, así como sobre sistemas complejos y no lineales.

Los resultados obtenidos al usar la eliminación de Gauss se pueden verificar al sustituirlos en las ecuaciones originales. No obstante, realizarlo no siempre representa una prueba confiable para sistemas mal condicionados. Por ello debe efectuarse alguna medida de la condición, como el determinante de un sistema escalado, si se tiene idea de que haya un error de redondeo. Dos opciones para disminuir el error de redondeo son el pivoteo parcial y el uso de un mayor número de cifras significativas en los cálculos. En el siguiente capítulo se regresará al tema de la condición del sistema cuando se analice la matriz inversa.

PROBLEMAS

9.1

a) Escriba en forma matricial el conjunto siguiente de ecuaciones:

$$8 = 6x_3 + 2x_2$$
$$2 - x_1 = x_3$$
$$5x_2 + 8x_1 = 13$$

b) Multiplique la matriz de coeficientes por su transpuesta; es decir $[A]\,[A]^T$.

9.2 Ciertas matrices están definidas como sigue

$$[A] = \begin{bmatrix} 4 & 7 \\ 1 & 2 \\ 5 & 6 \end{bmatrix} \qquad [B] = \begin{bmatrix} 4 & 3 & 7 \\ 1 & 2 & 7 \\ 2 & 0 & 4 \end{bmatrix}$$

$$\{C\} = \begin{Bmatrix} 3 \\ 6 \\ 1 \end{Bmatrix} \qquad [D] = \begin{bmatrix} 9 & 4 & 3 & -6 \\ 2 & -1 & 7 & 5 \end{bmatrix}$$

$$[E] = \begin{bmatrix} 1 & 5 & 8 \\ 7 & 2 & 3 \\ 4 & 0 & 6 \end{bmatrix}$$

$$[F] = \begin{bmatrix} 3 & 0 & 1 \\ 1 & 7 & 3 \end{bmatrix} \qquad \lfloor G \rfloor = \lfloor 7 \quad 6 \quad 4 \rfloor$$

En relación con estas matrices responda las preguntas siguientes:

a) ¿Cuáles son las dimensiones de las matrices?

b) Identifique las matrices cuadrada, columna y renglón.

c) ¿Cuáles son los valores de los elementos a_{12}, b_{23}, d_{32}, e_{22}, f_{12} y g_{12}?

d) Ejecute las operaciones siguientes:

1) $[E] + [B]$ 7) $[B] \times [A]$
2) $[A] + [F]$ 8) $[D]^T$
3) $[B] - [E]$ 9) $[A] \times \{C\}$
4) $7 \times [B]$ 10) $[I] \times [B]$
5) $[E] \times [B]$ 11) $[E]^T [E]$
6) $\{C\}^T$ 12) $\{C\}^T \{C\}$

9.3 Tres matrices se definen como

$$[A] = \begin{bmatrix} 1 & 6 \\ 3 & 10 \\ 7 & 4 \end{bmatrix} \quad [B] = \begin{bmatrix} 1 & 3 \\ 0.5 & 2 \end{bmatrix} \quad [C] = \begin{bmatrix} 2 & -2 \\ -3 & 1 \end{bmatrix}$$

a) Ejecute todas las multiplicaciones que sea posible calcular entre parejas de las matrices.
b) Utilice el método del cuadro PT3.2 para justificar por qué no se puede multiplicar a las demás parejas.
c) Emplee los resultados del inciso *a*) para ilustrar por qué es importante el orden de la multiplicación.

9.4 Use el método gráfico para resolver

$$4x_1 - 8x_2 = -24$$
$$-x_1 + 6x_2 = 34$$

Verifique sus resultados sustituyéndolos nuevamente en las ecuaciones.

9.5 Dado el sistema de ecuaciones

$$-1.1x_1 + 10x_2 = 120$$
$$-2x_1 + 17.4x_2 = 174$$

a) Resuélvalo gráficamente y compruebe el resultado con la sustitución en las ecuaciones.
b) Sobre la base de la solución gráfica, ¿qué se espera con respecto de la condición del sistema?
c) Calcule el determinante.
d) Resuelva por medio de la eliminación de incógnitas.

9.6 Para el sistema de ecuaciones que sigue

$$2x_2 + 5x_3 = 9$$
$$2x_1 + x_2 + x_3 = 9$$
$$3x_1 + x_2 = 10$$

a) Calcule el determinante.
b) Use la regla de Cramer para encontrar cuál es el valor de las x.
c) Sustituya el resultado en las ecuaciones originales para efectos de comprobación.

9.7 Dadas las ecuaciones

$$0.5x_1 - x_2 = -9.5$$
$$1.02x_1 - 2x_2 = -18.8$$

a) Resuelva en forma gráfica.
b) Calcule el determinante.
c) Con base en los incisos *a*) y *b*), ¿qué es de esperarse con respecto de la condición del sistema?
d) Resuelva por medio de la eliminación de incógnitas.
e) Resuelva otra vez, pero modifique ligeramente el elemento a_{11} a 0.52. Interprete sus resultados.

9.8 Dadas las ecuaciones siguientes

$$10x_1 + 2x_2 - x_3 = 27$$
$$-3x_1 - 6x_2 + 2x_3 = -61.5$$
$$x_1 + x_2 + 5x_3 = -21.5$$

a) Resuelva por eliminación de Gauss simple. Efectúe todos los pasos del cálculo.
b) Sustituya los resultados en las ecuaciones originales a fin de comprobar sus respuestas.

9.9 Use la eliminación de Gauss para resolver:

$$8x_1 + 2x_2 - 2x_3 = -2$$
$$10x_1 + 2x_2 + 4x_3 = 4$$
$$12x_1 + 2x_2 + 2x_3 = 6$$

Emplee pivoteo parcial y compruebe las respuestas sustituyéndolas en las ecuaciones originales.

9.10 Dado el sistema siguiente de ecuaciones

$$-3x_2 + 7x_3 = 2$$
$$x_1 + 2x_2 - x_3 = 3$$
$$5x_1 - 2x_2 = 2$$

a) Calcule el determinante.
b) Use la regla de Cramer para encontrar cuáles son los valores de las x.
c) Emplee la eliminación de Gauss con pivoteo parcial para obtener cuáles serían los valores de las x.
d) Sustituya sus resultados en las ecuaciones originales para efectos de comprobación.

9.11 Dadas las ecuaciones

$$2x_1 - 6x_2 - x_3 = -38$$
$$-3x_1 - x_2 + 7x_3 = -34$$
$$-8x_1 + x_2 - 2x_3 = -20$$

a) Resuelva por eliminación de Gauss con pivoteo parcial. Efectúe todos los pasos del cálculo.
b) Sustituya los resultados en las ecuaciones originales para comprobar sus respuestas.

9.12 Emplee la eliminación de Gauss-Jordan para resolver el sistema siguiente:

$$2x_1 + x_2 - x_3 = 1$$
$$5x_1 + 2x_2 + 2x_3 = -4$$
$$3x_1 + x_2 + x_3 = 5$$

No utilice pivoteo. Compruebe sus respuestas con la sustitución en las ecuaciones originales.

9.13 Resuelva el sistema:

$$x_1 + x_2 - x_3 = -3$$
$$6x_1 + 2x_2 + 2x_3 = 2$$
$$-3x_1 + 4x_2 + x_3 = 1$$

por medio de *a*) eliminación de Gauss simple, *b*) eliminación de Gauss con pivoteo parcial y *c*) método de Gauss-Jordan sin pivoteo parcial.

9.14 Lleve a cabo el mismo cálculo que en el ejemplo 9.11, pero use cinco paracaidistas con las características siguientes:

Paracaidista	Masa, kg	Coeficiente de arrastre, kg/s
1	55	10
2	75	12
3	60	15
4	75	16
5	90	10

Los paracaidistas tienen una velocidad de 9 m/s.

9.15 Resuelva el sistema

$$\begin{bmatrix} 3+2i & 4 \\ -i & 1 \end{bmatrix} \begin{Bmatrix} z_1 \\ z_2 \end{Bmatrix} = \begin{Bmatrix} 2+i \\ 3 \end{Bmatrix}$$

9.16 Desarrolle, depure y pruebe un programa en cualquier lenguaje de alto nivel o de macros de su predilección, para multiplicar dos matrices; es decir, $[X] = [Y]\,[Z]$, donde $[Y]$ es de orden m por n y $[Z]$ es de n por p. Pruebe el programa con el empleo de las matrices del problema 9.3.

9.17 Desarrolle, depure y pruebe un programa en cualquier lenguaje de alto nivel o de macros que prefiera, para generar la transpuesta de una matriz. Pruébelo con las matrices del problema 9.3.

9.18 Desarrolle, depure y pruebe un programa en el lenguaje de alto nivel o de macros que prefiera, para resolver un sistema de ecuaciones por medio de la eliminación de Gauss con pivoteo parcial. Base su programa en el pseudocódigo de la figura 9.6. Pruébelo con el uso del sistema siguiente (cuya respuesta es $x_1 = x_2 = x_3 = 1$),

$$x_1 + 2x_2 - x_3 = 2$$
$$5x_1 + 2x_2 + 2x_3 = 9$$
$$-3x_1 + 5x_2 - x_3 = 1$$

9.19 Tres masas están suspendidas verticalmente por una serie de resortes idénticos donde la masa 1 está en la parte superior y la masa 3 está en la parte inferior. Si $g = 9.81$ m/s^2, $m_1 = 2$ kg, $m_2 = 3$ kg, $m_3 = 2.5$ kg y las $k = 10$ kg/s^2, despeje los desplazamientos x.

9.20 Desarrolle, depure y pruebe un programa ya sea en lenguaje de alto nivel o en el macrolenguaje de su elección para resolver un sistema de n ecuaciones simultáneas lineales, con base en la sección 9.6. Pruebe el programa resolviendo el problema 7.12.

9.21 Recuerde, de la sección 8.2, que la química del agua expuesta al CO_2 atmosférico se puede determinar despejando las cinco incógnitas de las cinco ecuaciones simultáneas no lineales [ecuaciones (8.6) a (8.10)]: c_T, $[HCO_3^-]$, $[CO_3^{2-}]$, $[H^+]$ y $[OH^-]$. Usando los parámetros de la sección 8.2 y el programa desarrollado en el problema 9.20, resuelva para las condiciones de 1958, cuando la presión parcial del CO_2 era de 315 ppm. Use sus resultados para calcular el pH.

CAPÍTULO 10

Descomposición *LU* e inversión de matrices

En este capítulo se estudiará una clase de métodos de eliminación llamada técnicas de descomposición *LU*. El principal recurso de la descomposición *LU* es que el paso de la eliminación que toma mucho tiempo se puede formular de tal manera que involucre sólo operaciones con la matriz de coeficientes [*A*]. Por esto, es muy adecuado para aquellas situaciones donde se deben evaluar muchos vectores {*B*} del lado derecho para un solo valor de [*A*]. Aunque hay muchas formas de hacer esto, el análisis se enfocará en mostrar cómo el método de eliminación de Gauss se implementa como una descomposición *LU*.

Un motivo para introducir la descomposición *LU* es que proporciona un medio eficiente para calcular la matriz inversa. La inversa tiene muchas aplicaciones valiosas en la práctica de la ingeniería. Ésta ofrece también un medio para evaluar la condición de un sistema.

10.1 DESCOMPOSICIÓN *LU*

Como se describió en el capítulo anterior, la eliminación de Gauss sirve para resolver sistemas de ecuaciones algebraicas lineales,

$$[A]\{X\} = \{B\} \tag{10.1}$$

Aunque la eliminación de Gauss representa una forma satisfactoria para resolver tales sistemas, resulta ineficiente cuando deben resolverse ecuaciones con los mismos coeficientes [*A*], pero con diferentes constantes del lado derecho (las *b*).

Recuerde que la eliminación de Gauss implica dos pasos: eliminación hacia adelante y sustitución hacia atrás (figura 9.3). De ambas, el paso de eliminación hacia adelante es el que representa la mayor parte del trabajo computacional (recuerde la tabla 9.1). Esto es particularmente cierto para grandes sistemas de ecuaciones.

Los *métodos de descomposición LU* separan el tiempo usado en las eliminaciones para la matriz [*A*] de las manipulaciones en el lado derecho {*B*}. Una vez que [*A*] se ha “descompuesto”, los múltiples vectores del lado derecho {*B*} se pueden evaluar de manera eficiente.

El hecho de que la misma eliminación de Gauss se puede expresar como una descomposición *LU* es muy interesante. Antes de mostrar cómo se puede realizar esto, demos primero una demostración matemática de la estrategia de descomposición.

10.1.1 Revisión de la descomposición *LU*

De manera similar al caso de la eliminación de Gauss, la descomposición *LU* requiere de pivoteo para evitar la división entre cero. Sin embargo, para simplificar la siguiente descripción, abordaremos el tema del pivoteo después de que el planteamiento fundamental se haya elaborado. Además, la siguiente explicación se limita a un conjunto de tres ecuaciones simultáneas. Los resultados se pueden extender en forma directa a sistemas *n* dimensionales.

La ecuación (10.1) se reordena como

$$[A]\ \{X\} - \{B\} = 0 \tag{10.2}$$

Suponga que la ecuación (10.2) puede expresarse como un sistema triangular superior:

$$\begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} d_1 \\ d_2 \\ d_3 \end{Bmatrix} \tag{10.3}$$

Observe que esto es similar a la manipulación que ocurre en el primer paso de la eliminación de Gauss. Es decir, se utiliza la eliminación para reducir el sistema a una forma triangular superior. La ecuación (10.3) también se expresa en notación matricial y se reordena como

$$[U]\{X\} - \{D\} = 0 \tag{10.4}$$

Ahora, suponga que existe una matriz diagonal inferior con números 1 en la diagonal,

$$[L] = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \tag{10.5}$$

que tiene la propiedad de que cuando se premultiplica por la ecuación (10.4), el resultado es la ecuación (10.2). Es decir,

$$[L]\{[U]\{X\} - \{D\}\} = [A]\{X\} - \{B\} \tag{10.6}$$

Si esta ecuación se satisface, según las reglas de multiplicación entre matrices, se obtendrá

$$[L][U] = [A] \tag{10.7}$$

y

$$[L]\{D\} = \{B\} \tag{10.8}$$

Una estrategia de dos pasos (véase figura 10.1) para obtener soluciones se basa en las ecuaciones (10.4), (10.7) y (10.8):

1. *Paso de descomposición LU*. $[A]$ se factoriza o "descompone" en las matrices triangulares inferior $[L]$ y superior $[U]$.
2. *Paso de la sustitución*. $[L]$ y $[U]$ se usan para determinar una solución $\{X\}$ para un lado derecho $\{B\}$. Este paso, a su vez, se divide en dos. Primero, la ecuación (10.8) se usa para generar un vector intermedio $\{D\}$ mediante sustitución hacia adelante. Después, el resultado se sustituye en la ecuación (10.4), la que se resuelve por sustitución hacia atrás para $\{X\}$.

Ahora se mostrará cómo se puede llevar a cabo la eliminación de Gauss en esta forma.

10.1.2 Versión de la eliminación de Gauss usando la descomposición *LU*

Aunque a primera vista podría parecer que la eliminación de Gauss no está relacionada con la eliminación *LU*, aquélla puede usarse para descomponer $[A]$ en $[L]$ y $[U]$, lo cual se observa fácilmente para $[U]$, que es el resultado directo de la eliminación hacia adelante. Recuerde que en el paso correspondiente a esta eliminación se pretende reducir la matriz de coeficientes $[A]$ a la forma

$$[U] = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a'_{22} & a'_{23} \\ 0 & 0 & a''_{33} \end{bmatrix} \tag{10.9}$$

que es el formato triangular superior deseado.

Aunque quizá no sea muy clara, la matriz $[L]$ se produce durante este paso. Lo anterior se ilustra fácilmente con un sistema de tres ecuaciones,

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} b_1 \\ b_2 \\ b_3 \end{Bmatrix}$$

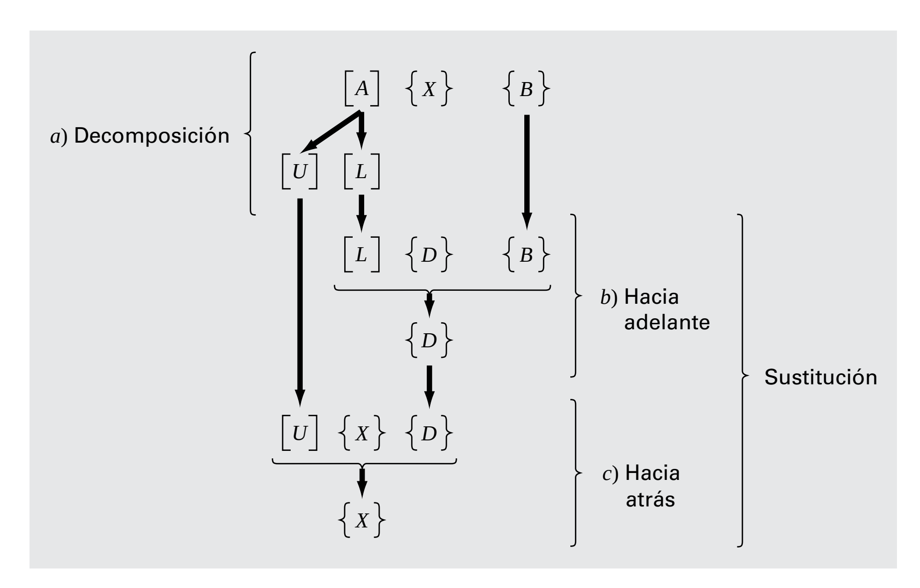


Figura 10.1
Pasos en la descomposición *LU*.

El primer paso en la eliminación de Gauss consiste en multiplicar el renglón 1 por el factor [recuerde la ecuación (9.13)]

$$f_{21} = \frac{a_{21}}{a_{11}}$$

y restar el resultado al segundo renglón para eliminar a_{21}. De forma similar, el renglón 1 se multiplica por

$$f_{31} = \frac{a_{31}}{a_{11}}$$

y el resultado se resta al tercer renglón para eliminar a_{31}. El paso final es multiplicar el segundo renglón modificado por

$$f_{32} = \frac{a'_{32}}{a'_{22}}$$

y restar el resultado al tercer renglón para eliminar a'_{32}.

Ahora suponga que realizamos todas esas operaciones sólo en la matriz [A]. Resulta claro que si no se quiere modificar la ecuación, se tiene que hacer lo mismo con el lado derecho {B}. Pero no existe ninguna razón para realizar las operaciones en forma simultánea. Se podrían conservar las f y después manipular {B}.

¿Dónde se guardan los factores f_{21}, f_{31} y f_{32}? Recuerde que la idea principal de la eliminación fue crear ceros en a_{21}, a_{31} y a_{32}. Entonces, se puede guardar f_{21} en a_{21}, f_{31} en a_{31} y f_{32} en a_{32}. Después de la eliminación la matriz [A], por lo tanto, se describe como

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ f_{21} & a'_{22} & a'_{23} \\ f_{31} & f_{32} & a''_{33} \end{bmatrix} \tag{10.10}$$

De hecho, esta matriz representa un almacenamiento eficiente de la descomposición *LU* de [A],

$$[A] \rightarrow [L][U] \tag{10.11}$$

donde

$$[U] = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a'_{22} & a'_{23} \\ 0 & 0 & a''_{33} \end{bmatrix}$$

y

$$[L] = \begin{bmatrix} 1 & 0 & 0 \\ f_{21} & 1 & 0 \\ f_{31} & f_{32} & 1 \end{bmatrix}$$

El siguiente ejemplo confirma que $[A] = [L][U]$.

EJEMPLO 10.1 Descomposición *LU* con eliminación de Gauss

Planteamiento del problema Obtenga una descomposición *LU* basándose en la eliminación de Gauss que se realizó en el ejemplo 9.5.

Solución En el ejemplo 9.5, se resolvió la matriz

$$[A] = \begin{bmatrix} 3 & -0.1 & -0.2 \\ 0.1 & 7 & -0.3 \\ 0.3 & -0.2 & 10 \end{bmatrix}$$

Después de la eliminación hacia adelante, se obtuvo la siguiente matriz triangular superior:

$$[U] = \begin{bmatrix} 3 & -0.1 & -0.2 \\ 0 & 7.00333 & -0.293333 \\ 0 & 0 & 10.0120 \end{bmatrix}$$

Los factores empleados para obtener la matriz triangular superior se pueden colocar en una matriz triangular inferior. Los elementos a_{21} y a_{31} se eliminaron al usar los factores

$$f_{21} = \frac{0.1}{3} = 0.03333333 \qquad f_{31} = \frac{0.3}{3} = 0.1000000$$

y el elemento a'_{32} se elimina al usar el factor

$$f_{32} = \frac{-0.19}{7.00333} = -0.0271300$$

Así, la matriz triangular inferior es

$$[L] = \begin{bmatrix} 1 & 0 & 0 \\ 0.0333333 & 1 & 0 \\ 0.100000 & -0.0271300 & 1 \end{bmatrix}$$

En consecuencia, la descomposición *LU* es

$$[A] = [L][U] = \begin{bmatrix} 1 & 0 & 0 \\ 0.0333333 & 1 & 0 \\ 0.100000 & -0.0271300 & 1 \end{bmatrix} \begin{bmatrix} 3 & -0.1 & -0.2 \\ 0 & 7.00333 & -0.293333 \\ 0 & 0 & 10.0120 \end{bmatrix}$$

Este resultado se verifica al realizar la multiplicación de $[L][U]$ que da

$$[L][U] = \begin{bmatrix} 3 & -0.1 & -0.2 \\ 0.0999999 & 7 & -0.3 \\ 0.3 & -0.2 & 9.99996 \end{bmatrix}$$

donde las pequeñas diferencias se deben a errores de redondeo.

El siguiente es el pseudocódigo de una subrutina para realizar la fase de descomposición:

```
SUB Decompose (a, n)
  DOFOR k = 1, n - 1
    DOFOR i = k + 1, n
      factor = a_i,k/a_k,k
      a_i,k = factor
      DOFOR j = k + 1, n
        a_i,j = a_i,j - factor * a_k,j
      END DO
    END DO
  END DO
END Decompose
```

Observe que este algoritmo es "simple" en el sentido de que no se incluye el pivoteo. Esta característica se agregará más tarde cuando se desarrolle el algoritmo completo para la descomposición *LU*.

Después de descomponer la matriz, se puede generar una solución para un vector particular $\{B\}$. Esto se lleva a cabo en dos pasos. Primero, se realiza un paso de sustitución hacia adelante al resolver la ecuación (10.8) para $\{D\}$. Es importante notar que esto sólo se refiere a la realización de las operaciones de la eliminación en $\{B\}$. De esta forma, al final del procedimiento, el lado derecho estará en el mismo estado que si se hubiesen realizado las operaciones hacia adelante sobre $[A]$ y $\{B\}$ en forma simultánea.

El paso de la sustitución hacia adelante se representa en forma concisa como

$$d_i = b_i - \sum_{j=1}^{i-1} a_{ij} d_j \qquad \text{para } i = 2, 3, \ldots, n \tag{10.12}$$

En el segundo paso, entonces, tan sólo se realiza la sustitución hacia atrás, como en la ecuación (10.4). Otra vez, es importante reconocer que este paso es idéntico al de la fase de sustitución hacia atrás, en la eliminación de Gauss convencional. Así, de manera similar a las ecuaciones (9.16) y (9.17), el paso de la sustitución hacia atrás se representa en forma concisa como

$$x_n = d_n / a_{nn} \tag{10.13}$$

$$x_i = \frac{d_i - \sum_{j=i+1}^{n} a_{ij} x_j}{a_{ii}} \qquad \text{para } i = n-1, n-2, \ldots, 1 \tag{10.14}$$

EJEMPLO 10.2 Pasos en la sustitución

Planteamiento del problema Termine el problema que se inició en el ejemplo 10.1 para generar la solución final con eliminación hacia adelante y sustitución hacia atrás.

Solución Como se estableció antes, la intención de la sustitución hacia adelante es aplicar las operaciones de eliminación al vector $\{B\}$, previamente aplicadas a $[A]$. Recuerde que el sistema resuelto en el ejemplo 9.5 fue

$$\begin{bmatrix} 3 & -0.1 & -0.2 \\ 0.1 & 7 & -0.3 \\ 0.3 & -0.2 & 10 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 7.85 \\ -19.3 \\ 71.4 \end{Bmatrix}$$

y que la fase de eliminación hacia adelante del método de eliminación convencional de Gauss dio como resultado

$$\begin{bmatrix} 3 & -0.1 & -0.2 \\ 0 & 7.00333 & -0.293333 \\ 0 & 0 & 10.0120 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 7.85 \\ -19.5617 \\ 70.0843 \end{Bmatrix} \tag{E10.2.1}$$

La fase de la sustitución hacia adelante se realiza aplicando la ecuación (10.7) a nuestro problema,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.0333333 & 1 & 0 \\ 0.100000 & -0.0271300 & 1 \end{bmatrix} \begin{Bmatrix} d_1 \\ d_2 \\ d_3 \end{Bmatrix} = \begin{Bmatrix} 7.85 \\ -19.3 \\ 71.4 \end{Bmatrix}$$

o, despejando por multiplicación del lado izquierdo,

$$\begin{aligned} d_1 &= 7.85 \\ 0.0333333 d_1 + d_2 &= -19.3 \\ 0.1 d_1 - 0.02713 d_2 + d_3 &= 71.4 \end{aligned}$$

Se resuelve la primera ecuación para d_1,

$$d_1 = 7.85$$

la cual se sustituye en la segunda ecuación y se resuelve para

$$d_2 = -19.3 - 0.0333333(7.85) = -19.5617$$

Ambas, d_1 y d_2, se sustituyen en la tercera ecuación para

$$d_3 = 71.4 - 0.1(7.85) + 0.02713(-19.5617) = 70.0843$$

Así,

$$\{D\} = \begin{Bmatrix} 7.85 \\ -19.5617 \\ 70.0843 \end{Bmatrix}$$

que es idéntica al lado derecho de la ecuación (E10.2.l).

Este resultado se sustituye, entonces, en la ecuación (10.4), $[U]\{X\} = \{D\}$, para obtener

$$\begin{bmatrix} 3 & -0.1 & -0.2 \\ 0 & 7.00333 & -0.293333 \\ 0 & 0 & 10.0120 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 7.85 \\ -19.5617 \\ 70.0843 \end{Bmatrix}$$

que se resuelve por sustitución hacia atrás (véase el ejemplo 9.5 para más detalles) para obtener la solución final,

$$\{X\} = \begin{Bmatrix} 3 \\ -2.5 \\ 7.00003 \end{Bmatrix}$$

El siguiente es el pseudocódigo de una subrutina para implementar ambas fases de sustitución:

```
SUB Substitute (a, n, b, x)
  'sustitución hacia adelante
  DOFOR i = 2, n
    sum = b_i
    DOFOR j = 1, i - 1
      sum = sum - a_i,j * b_j
    END DO
    b_i = sum
  END DO
  'sustitución hacia atrás
  x_n = b_n / a_n,n
  DOFOR i = n - 1, 1, -1
    sum = 0
    DOFOR j = i + 1, n
      sum = sum + a_i,j * x_j
    END DO
    x_i = (b_i - sum)/a_i,i
  END DO
END Substitute
```

El algoritmo de descomposición *LU* requiere los mismos flops de multiplicación/división totales que la eliminación de Gauss. La única diferencia es que se aplica un menor trabajo en la fase de descomposición, debido a que las operaciones no se aplican al lado derecho. De esta forma, el número de flops de multiplicación/división en la fase de descomposición se calculan así:

$$\frac{n^3}{3} - \frac{n}{3} \xrightarrow[\text{aumenta}]{\text{conforme } n} \frac{n^3}{3} + O(n) \tag{10.15}$$

Por lo contrario, la fase de sustitución requiere de un mayor trabajo. Así, el número de flops para la sustitución hacia adelante y hacia atrás es n^2. El trabajo total es, por lo tanto, idéntico al de la eliminación de Gauss

$$\frac{n^3}{3} - \frac{n}{3} + n^2 \xrightarrow[\text{aumenta}]{\text{conforme } n} \frac{n^3}{3} + O(n^2) \tag{10.16}$$

10.1.3 Algoritmo para la descomposición *LU*

En la figura 10.2 se presenta un algoritmo que implementa la descomposición *LU* con eliminación de Gauss. Vale la pena mencionar cuatro características de este algoritmo:

- Los factores generados durante la fase de eliminación se guardan en la parte inferior de la matriz. Esto puede hacerse debido a que de cualquier manera éstos se convierten en ceros y no son necesarios en la solución final. Este almacenamiento ahorra espacio.
- El algoritmo lleva cuenta del pivoteo al usar un vector de orden O. Esto acelera notablemente el algoritmo, ya que sólo se pivotea el vector (y no todo el renglón).
- Las ecuaciones no están escaladas, pero se usan valores escalados de los elementos para determinar si se va a usar el pivoteo.
- El término de la diagonal se verifica durante la fase de pivoteo para detectar ocurrencias cercanas a cero con el propósito de advertir al usuario respecto de sistemas singulares. Si baja de un valor er = –1, entonces se ha detectado una matriz singular y se debe terminar el cálculo. El usuario le da a un parámetro *tol* un valor pequeño, para detectar ocurrencias cercanas a cero.

```
SUB Ludecomp (a, b, n, tol, x, er)
  DIM on, sn
  er = 0
  CALL Decompose(a, n, tol, o, s, er)
  IF er <> -1 THEN
     CALL Substitute(a, o, n, b, x)
  END IF
END Ludecomp

SUB Decompose (a, n, tol, o, s, er)
  DOFOR i = 1, n
    oi = i
    si = ABS(ai,1)
    DOFOR j = 2, n
       IF ABS(ai,j)>si THEN si = ABS(ai,j)
    END DO
  END DO
  DOFOR k = 1, n - 1
    CALL Pivot(a, o, s, n, k)
    IF ABS(ao(k),k /so(k)) < tol THEN
       er = -1
       PRINT ao(k),k/so(k)
       EXIT DO
     END IF
     DOFOR i = k + 1, n
       factor = ao(i),k /ao(k),k
       ao(i),k = factor
       DOFOR j = k + 1, n
         ao(i),j = ao(i),j - factor * ao(k),j
        END DO
     END DO
  END DO
  IF ABS(ao(k),k/so(k)) < tol THEN
     er = -1
     PRINT ao(k),k/so(k)
  END IF
END Decompose

SUB Pivot(a, o, s, n, k)
  p = k
  big = ABS(ao(k),k /so(k))
  DOFOR ii = k + 1, n
    dummy = ABS(ao(ii),k /so(ii))
    IF dummy > big THEN
       big = dummy
       p = ii
    END IF
  END DO
  dummy = op
  op = ok
  ok = dummy
END Pivot

SUB Substitute (a, o, n, b, x)
  DOFOR i = 2, n
    sum = bo(i)
    DOFOR j = 1, i -1
      sum = sum - ao(i),j * bo(j)
     END DO
    bo(i) = sum
  END DO
  xn = bo(n) /ao(n),n
  DOFOR i = n - 1, 1, -1
    sum = 0
    DOFOR j = i + 1, n
      sum + sum + ao(i),j * xj
    END DO
    xi = (bo(i) - sum)/ao(i),i
  END DO
END Substitute
```



Figura 10.2
Pseudocódigo para un algoritmo de descomposición *LU*.

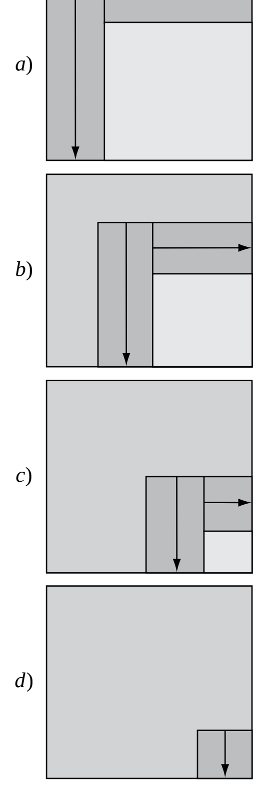


Figura 10.3
Un esquema que muestra las evaluaciones implicadas en la descomposición *LU* de Crout.

10.1.4 Descomposición Crout

Observe que en la descomposición *LU* con la eliminación de Gauss, la matriz [*L*] tiene números 1 en la diagonal. Formalmente, a esto se le denomina *descomposición* o *factorización de Doolittle*. Un método alternativo usa una matriz [*U*] con números 1 sobre la diagonal. Esto se conoce como *descomposición Crout*. Aunque hay algunas diferencias entre estos métodos, su funcionamiento es comparable (Atkinson, 1978; Ralston y Rabinowitz, 1978).

El método de descomposición de Crout genera [*U*] y [*L*] barriendo las columnas y los renglones de la matriz, como se ilustra en la figura 10.3. La descomposición de Crout se puede implementar mediante la siguiente serie concisa de fórmulas:

$$l_{i,1} = a_{i,1} \qquad \text{para } i = 1, 2, \ldots, n \tag{10.17}$$

$$u_{1j} = \frac{a_{1j}}{l_{11}} \qquad \text{para } j = 2, 3, \ldots, n \tag{10.18}$$

Para $j = 2, 3, \ldots, n - 1$

$$l_{ij} = a_{ij} - \sum_{k=1}^{j-1} l_{ik}u_{kj} \qquad \text{para } i = j, j+1, \ldots, n \tag{10.19}$$

$$u_{jk} = \frac{a_{jk} - \sum_{i=1}^{j-1} l_{ji}u_{ik}}{l_{jj}} \qquad \text{para } k = j+1, j+2, \ldots, n \tag{10.20}$$

y

$$l_{nn} = a_{nn} - \sum_{k=1}^{n-1} l_{nk}u_{kn} \tag{10.21}$$

Además de que consiste de pocos ciclos, el método anterior también tiene la ventaja de economizar espacio de almacenamiento. No hay necesidad de guardar los números 1 que están en la diagonal de [U] o los números cero de [L] o [U], ya que se dan en el método. En consecuencia, los valores de [U] se pueden guardar en el espacio de los ceros de [L]. Además, mediante un cuidadoso examen de lo anterior, queda claro que después de que un elemento de [A] se emplea una vez, nunca vuelve a utilizarse. Por lo tanto, conforme se va calculando cada elemento de [L] y [U], se puede sustituir por el elemento correspondiente de [A] (como se designó por sus subíndices).

El pseudocódigo para realizar esto se presenta en la figura 10.4. Observe que la ecuación (10.17) no está incluida en el pseudocódigo, porque la primera columna de [L] ya se guardó en [A]. De otra forma, el algoritmo sigue, en forma directa, de las ecuaciones (10.18) a la (10.2l).

```
DOFOR j = 2, n
  a1,j = a1,j/a1,1
END DO
DOFOR j = 2, n - 1
  DOFOR i = j, n
    sum = 0
    DOFOR k = 1, j - 1
      sum = sum + ai,k · ak,j
    END DO
    ai,j = ai,j - sum
  END DO
  DOFOR k = j + 1, n
    sum = 0
    DOFOR i = 1, j - 1
      sum = sum + aj,i · ai,k
    END DO
    aj,k = (aj,k - sum)/aj,j
  END DO
END DO
sum = 0
DOFOR k = 1, n - 1
  sum = sum + an,k · ak,n
END DO
an,n = an,n - sum
```

Figura 10.4
Pseudocódigo para el algoritmo de la descomposición *LU* de Crout.

10.2 LA MATRIZ INVERSA

En el estudio de las operaciones con matrices (sección PT3.2.2), vimos que si una matriz [A] es cuadrada, existe otra matriz $[A]^{-1}$, conocida como la inversa de [A], para la cual [ecuación (PT3.3)]

$$[A][A]^{-1} = [A]^{-1}[A] = [I]$$

Ahora se enfocará el análisis hacia el modo en que la matriz inversa se calcula numéricamente. Después se explorará cómo se utiliza para el diseño en ingeniería.

10.2.1 Cálculo de la inversa

La inversa se puede calcular en forma de columna por columna, generando soluciones con vectores unitarios como las constantes del lado derecho. Por ejemplo, si la constante del lado derecho de la ecuación tienen un número 1 en la primera posición, y ceros en las otras,

$$\{b\} = \begin{Bmatrix} 1 \\ 0 \\ 0 \end{Bmatrix}$$

la solución resultante será la primera columna de la matriz inversa. En forma similar, si se emplea un vector unitario que tiene un número 1 en el segundo renglón

$$\{b\} = \begin{Bmatrix} 0 \\ 1 \\ 0 \end{Bmatrix}$$

el resultado será la segunda columna de la matriz inversa.

La mejor forma de realizar un cálculo como éste es con el algoritmo de descomposición *LU*, descrito al inicio de este capítulo. Recuerde que una de las ventajas más importantes de la descomposición *LU* es que proporciona un medio eficiente para evaluar diversos vectores del lado derecho. Por lo tanto, resulta ideal para evaluar los vectores unitarios requeridos en el cálculo de la inversa.

EJEMPLO 10.3 Inversión de matrices

Planteamiento del problema Emplee la descomposición *LU* para determinar la matriz inversa del sistema del ejemplo 10.2.

$$[A]=\begin{bmatrix} 3 & -0.1 & -0.2 \\ 0.1 & 7 & -0.3 \\ 0.3 & -0.2 & 10 \end{bmatrix}$$

Recuerde que la descomposición dio como resultado las siguientes matrices triangulares inferior y superior:

$$[U]=\begin{bmatrix} 3 & -0.1 & -0.2 \\ 0 & 7.00333 & -0.293333 \\ 0 & 0 & 10.0120 \end{bmatrix} \quad [L]=\begin{bmatrix} 1 & 0 & 0 \\ 0.0333333 & 1 & 0 \\ 0.100000 & -0.0271300 & 1 \end{bmatrix}$$

Solución La primera columna de la matriz inversa puede determinarse al efectuar el procedimiento de solución por sustitución hacia adelante, con un vector unitario (con el número 1 en el primer renglón) como el vector del lado derecho. Así, de la ecuación (10.8), el sistema diagonal inferior es

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.0333333 & 1 & 0 \\ 0.100000 & -0.0271300 & 1 \end{bmatrix} \begin{Bmatrix} d_1 \\ d_2 \\ d_3 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 0 \\ 0 \end{Bmatrix}$$

de donde, por sustitución hacia adelante, se obtiene $\{D\}^T = [1 \quad -0.03333 \quad -0.1009]$. Este vector se utiliza como el lado derecho de la ecuación (10.3),

$$\begin{bmatrix} 3 & -0.1 & -0.2 \\ 0 & 7.00333 & -0.293333 \\ 0 & 0 & 10.0120 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 1 \\ -0.03333 \\ -0.1009 \end{Bmatrix}$$

de donde, por sustitución hacia atrás, se obtiene $\{X\}^T = [0.33249 \quad -0.00518 \quad -0.01008]$, que es la primera columna de la matriz,

$$[A]^{-1}=\begin{bmatrix} 0.33249 & 0 & 0 \\ -0.00518 & 0 & 0 \\ -0.01008 & 0 & 0 \end{bmatrix}$$

Para determinar la segunda columna, la ecuación (10.8) se formula como

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.0333333 & 1 & 0 \\ 0.100000 & -0.0271300 & 1 \end{bmatrix} \begin{Bmatrix} d_1 \\ d_2 \\ d_3 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 1 \\ 0 \end{Bmatrix}$$

De donde se puede obtener $\{D\}$, y los resultados se usan con la ecuación (10.3) para determinar $\{X\}^T = [0.004944 \quad 0.142903 \quad 0.00271]$, que es la segunda columna de la matriz,

$$[A]^{-1} = \begin{bmatrix} 0.33249 & 0.004944 & 0 \\ -0.00518 & 0.142903 & 0 \\ -0.01008 & 0.00271 & 0 \end{bmatrix}$$

Por último, los procedimientos de sustitución hacia adelante y de sustitución hacia atrás pueden usarse con $\{B\}^T = \lfloor 0 \;\; 0 \;\; 1 \rfloor$, para obtener $\{X\}^T = [0.006798 \quad 0.004183 \quad 0.09988]$, que es la columna final de la matriz,

$$[A]^{-1} = \begin{bmatrix} 0.33249 & 0.004944 & 0.006798 \\ -0.00518 & 0.142903 & 0.004183 \\ -0.01008 & 0.00271 & 0.09988 \end{bmatrix}$$

La validez de este resultado se comprueba al verificar que $[A][A]^{-1} = [I]$.

El pseudocódigo para generar la matriz inversa se muestra en la figura 10.5. Observe cómo se llama a la subrutina de descomposición de la figura 10.2, para realizar la descomposición, y después se genera la inversa llamando repetidamente el algoritmo de sustitución con vectores unitarios.

El trabajo requerido para este algoritmo se calcula fácilmente como

$$\underset{\text{descomposición}}{\frac{n^3}{3} - \frac{n}{3}} \quad + \quad \underset{n \,\times\, \text{sustituciones}}{n(n^2)} \quad = \frac{4n^3}{3} - \frac{n}{3} \tag{10.22}$$

donde, de acuerdo con la sección 10.1.2, la descomposición está definida por la ecuación (10.15) y el trabajo necesario en cada evaluación del lado derecho requiere n^2 flops de multiplicación/ división.

10.2.2 Cálculos estímulo-respuesta

Como se vio en la sección PT3.1.2, muchos de los sistemas de ecuaciones lineales usados en la práctica de la ingeniería se obtienen de las leyes de la conservación. La expresión matemática de dichas leyes es algún tipo de ecuación de balance que asegura que una propiedad específica se conserve (masa, fuerza, calor, *momentum* u otra). En un balance de fuerzas de una estructura, las propiedades pueden ser los componentes horizontal o vertical de las fuerzas que actúan sobre cada nodo de la estructura (véase la sección 12.2). En un balance de masa, las propiedades pueden ser la masa en cada reactor de un proceso químico (véase la sección 12.1). Se tendrán ejemplos similares en otros campos de la ingeniería.

Al tenerse una ecuación de balance para cada parte del sistema, da como resultado un conjunto de ecuaciones que definen el comportamiento de las propiedades en todo el sistema. Estas ecuaciones se interrelacionan, ya que cada ecuación puede tener una o más de las variables de las otras ecuaciones. En muchos casos, estos sistemas son lineales y, por lo tanto, de la forma que se trata en este capítulo:

$$[A]\{X\} = \{B\} \tag{10.23}$$

Ahora bien, para las ecuaciones de balance, los términos de la ecuación (10.23) tienen una interpretación física definida. Por ejemplo, los elementos de $\{X\}$ son los valores de la propiedad que se balanceará en cada parte del sistema. En el balance de fuerzas de una estructura, representan las fuerzas vertical y horizontal en cada miembro. En el balance de masa, los elementos de $\{X\}$ son las masas de sustancias químicas en cada reactor. En cualquier caso, representan la respuesta o estado del sistema, que se está tratando de determinar.

```
CALL Decompose (a, n, tol, o, s, er)
IF er = 0 THEN
   DOFOR i = 1, n
     DOFOR j = 1, n
       IF i = j THEN
          b(j) = 1
       ELSE
          b(j) = 0
       END IF
     END DO
     Call Substitute (a, o, n, b, x)
     DOFOR j = 1, n
       ai(j, i) = x(j)
     END DO
   END DO
   salida ai, si lo desea
ELSE
   PRINT “sistema mal condicionado”
END IF
```

Figura 10.5
Programa principal que usa algunos de los subprogramas de la figura 10.2 para generar una matriz inversa.

El vector del lado derecho $\{B\}$ contiene los elementos del balance que son independientes del comportamiento del sistema (es decir, son constantes). Como tales, representan las fuerzas externas o los estímulos que rigen al sistema.

Finalmente, la matriz de coeficientes $[A]$ contiene los parámetros que expresan cómo interactúan las partes del sistema. En consecuencia, la ecuación (10.23) se puede expresar como:

$$[\text{interacciones}]\{\text{respuesta}\} = \{\text{estímulos}\}$$

Así, la ecuación (10.23) puede verse como una expresión del modelo matemático fundamental que se formuló anteriormente como una sola ecuación en el capítulo 1 [recuerde la ecuación (1.1)]. Ahora se percibe que la ecuación (10.23) representa una versión para sistemas interrelacionados con diversas variables dependientes $\{X\}$.

Como ya hemos visto en este capítulo y en el anterior, existen varias formas de resolver la ecuación (10.23). Sin embargo, usando la matriz inversa se obtiene un resultado particularmente interesante. La solución formal se expresa como

$$\{X\} = [A]^{-1}\{B\}$$

o (recordando la definición de la multiplicación matricial del cuadro PT3.2)

$$x_1 = a_{11}^{-1} b_1 + a_{12}^{-1} b_2 + a_{13}^{-1} b_3$$
$$x_2 = a_{21}^{-1} b_1 + a_{22}^{-1} b_2 + a_{23}^{-1} b_3$$
$$x_3 = a_{31}^{-1} b_1 + a_{32}^{-1} b_2 + a_{33}^{-1} b_3$$

De esta forma, se ha encontrado que la misma matriz inversa, además de ofrecer una solución, tiene propiedades extremadamente útiles. Es decir, cada uno de sus elementos representa la respuesta de una sola parte del sistema a un estímulo unitario de cualquier otra parte de dicho sistema.

Observe que estas formulaciones son lineales y, por lo tanto, se satisfacen la superposición y la proporcionalidad. La *superposición* significa que si un sistema está sujeto a varios estímulos (las b), las respuestas se pueden calcular individualmente y los resultados se suman para obtener la respuesta total. La *proporcionalidad* significa que al multiplicar los estímulos por una cantidad el resultado es la respuesta a esos estímulos multiplicada por la misma cantidad. Así, el coeficiente a_{11}^{-1} es una constante de proporcionalidad que da el valor de x_1 correspondiente a una cantidad unitaria b_1. Este resultado es independiente de los efectos de b_2 y b_3 sobre x_1, los cuales se reflejan en los coeficientes a_{12}^{-1} y a_{13}^{-1}, respectivamente. Por lo tanto, se llega a la conclusión general de que el elemento a_{ij}^{-1} de la matriz inversa representa el valor de x_1 debido a la cantidad unitaria b_j. Usando el ejemplo de la estructura, el elemento a_{ij}^{-1} de la matriz inversa representaría la fuerza en el miembro i debida a una fuerza unitaria externa en el nodo j. Incluso para sistemas pequeños, dicho comportamiento de interacciones estímulo-respuesta individuales podría no ser intuitivamente obvio. Como tal, la matriz inversa ofrece una poderosa técnica para comprender las interrelaciones entre las partes componentes de sistemas complicados. Este poder se demostrará en las secciones 12.1 y 12.2.

10.3 ANÁLISIS DEL ERROR Y CONDICIÓN DEL SISTEMA

Además de sus aplicaciones a la ingeniería, la inversa también proporciona un medio para determinar si los sistemas están mal condicionados. Están disponibles tres métodos para este propósito:

1. Escalar la matriz de coeficientes $[A]$, de manera que el elemento más grande en cada renglón sea 1. Se invierte la matriz escalada, y si existen elementos de $[A]^{-1}$ que sean varios órdenes de magnitud mayores que uno, es posible que el sistema esté mal condicionado (véase el cuadro 10.1).
2. Multiplicar la inversa por la matriz de coeficientes original y estimar si el resultado es lo suficientemente cercano a la matriz identidad. Si no es así, esto indica que el sistema está mal condicionado.
3. Invertir la matriz inversa y estimar si el resultado está lo suficientemente cercano a la matriz de coeficientes original. Si no es así, esto de nueva cuenta indica que el sistema está mal condicionado.

Cuadro 10.1 Interpretación de los elementos de la matriz inversa como una medida de mal condicionamiento

Un método para determinar la condición de un sistema consiste en escalar $[A]$ de tal forma que el elemento mayor en cada renglón sea 1 y después calcular $[A]^{-1}$. Si los elementos de $[A]^{-1}$ son varios órdenes de magnitud mayores que los elementos de la matriz escalada original, es probable que el sistema esté mal condicionado.

Se puede obtener cierto conocimiento con este método al recordar que una forma de verificar si una solución aproximada $\{X\}$ es aceptable, es sustituyéndola en las ecuaciones originales y observar si resultan las constantes originales del lado derecho. Esto equivale a

$$\{R\} = \{B\} - [A]\{\tilde{X}\} \tag{C10.1.1}$$

donde $\{R\}$ es el residuo entre las constantes del lado derecho y los valores calculados con la solución $\{\tilde{X}\}$. Si $\{R\}$ es pequeño, se concluye que los valores de $\{\tilde{X}\}$ son adecuados. Suponiendo que $\{X\}$ es la solución exacta que da un residuo cero, entonces

$$\{0\} = \{B\} - [A]\{X\} \tag{C10.1.2}$$

Restando la ecuación (C10.1.2) de (C10.1.1) resulta

$$\{R\} = [A]\left\{\{X\} - \{\tilde{X}\}\right\}$$

Multiplicando ambos lados de esta ecuación por $[A]^{-1}$ se obtiene

$$\{X\} - \{\tilde{X}\} = [A]^{-1}\{R\}$$

Este resultado indica por qué la verificación de una solución por sustitución puede ser engañosa. Para casos donde los elementos de $[A]^{-1}$ son grandes, una pequeña discrepancia en el residuo $\{R\}$ del lado derecho, puede corresponder a un gran error $\{X\} - \{\tilde{X}\}$ en el valor calculado de las incógnitas. En otras palabras, un residuo pequeño no garantiza una solución exacta. Aunque, puede concluirse que si el elemento mayor de $[A]^{-1}$ es de un orden de magnitud unitaria, se puede considerar que el sistema está bien condicionado. De modo contrario, si $[A]^{-1}$ contiene elementos mucho más grandes que la unidad se concluye que el sistema está mal condicionado.

Aunque estos métodos llegan a indicar un mal condicionamiento, sería preferible obtener un solo número (al igual que el número de condición de la sección 4.2.3) que sirviera como un indicador del problema. Los intentos que se han hecho para formular tal número de condición matricial están basados en el concepto matemático de la norma.

10.3.1 Normas vectoriales y matriciales

Una *norma* es una función que toma valores reales y que proporciona una medida del tamaño o "longitud" de entidades matemáticas multicomponentes, como los vectores y las matrices (véase cuadro 10.2).

Un ejemplo simple es un vector en el espacio euclidiano tridimensional (figura 10.6) que se representa como

$$\lfloor F \rfloor = \lfloor a \;\; b \;\; c \rfloor$$

donde a, b y c son las distancias a lo largo de los ejes x, y y z, respectivamente. La longitud de este vector [esto es, la distancia de la coordenada (0, 0, 0) a (a, b, c)] se calcula simplemente como

$$\|F\|_e = \sqrt{a^2 + b^2 + c^2}$$

donde la nomenclatura $\|F\|_e$ indica que a esta longitud se refiere a la norma euclidiana de $[F]$.

En forma similar, para un vector n dimensional $\lfloor X \rfloor = \lfloor x_1 \;\; x_2 \;\; \cdots \;\; x_n \rfloor$, una norma euclidiana se calcularía como

$$\|X\|_e = \sqrt{\sum_{i=1}^{n} x_i^2}$$

El concepto puede extenderse además a una matriz $[A]$, de la siguiente manera

$$\|A\|_e = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{i,j}^2} \tag{10.24}$$

a la cual se le da un nombre especial (la *norma de Frobenius*). De la misma manera como las normas de vectores, proporciona un valor único para cuantificar el "tamaño" de $[A]$.

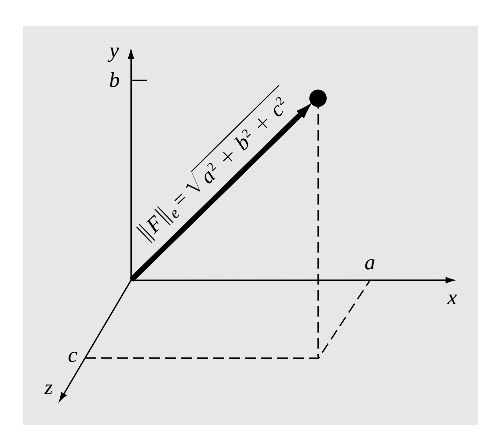


Figura 10.6
Representación gráfica de un vector $\lfloor F \rfloor = [a \;\; b \;\; c]$ en el espacio euclidiano.

Debe notarse que hay alternativas para las normas euclidiana y de Frobenius (véase cuadro 10.2). Por ejemplo, *la norma vector uniforme* se define como

$$\|X\|_\infty = \max_{1 \le i \le n} |x_i|$$

Es decir, el elemento con el mayor valor absoluto se toma como la medida del tamaño del vector. En forma similar, una *norma matricial uniforme* o *norma renglón-suma* se define como

$$\|A\|_\infty = \max_{1 \le i \le n} \sum_{j=1}^{n} |a_{ij}| \tag{10.25}$$

En este caso, se calcula la suma del valor absoluto de los elementos por cada renglón, y la mayor de éstas se toma como la norma.

Aunque hay ventajas teóricas para el uso de ciertas normas, la elección está influenciada por consideraciones prácticas. Por ejemplo, la norma renglón-uniforme es ampliamente usada por la facilidad con que se calcula, y porque usualmente proporciona una medida adecuada del tamaño de la matriz.

10.3.2 Número de condición de una matriz

Ahora que se ha presentado el concepto de norma, se puede usar para definir

$$\text{Cond } [A] = \|A\| \cdot \|A^{-1}\| \tag{10.26}$$

donde Cond [*A*] se llama *número de condición de una matriz*. Observe que para una matriz [*A*], este número será mayor o igual a 1. Se puede mostrar (Ralston y Rabinowitz, 1978; Gerald y Wheatley, 2004) que

Cuadro 10.2 Normas matriciales

Como se vio en esta sección, las normas euclidianas se emplean para cuantificar el tamaño de un vector,

$$\|X\|_e = \sqrt{\sum_{i=1}^{n} x_i^2}$$

o de una matriz,

$$\|A\|_e = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j}^2}$$

Para vectores, existen alternativas llamadas normas *p* que se representan generalmente por

$$\|X\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

Puede observarse que la norma euclidiana y la norma 2, $\|X\|_2$, son idénticas para vectores.

Otros ejemplos importantes son

$$\|X\|_1 = \sum_{i=1}^{n} |x_i|$$

que representa la norma como la suma de los valores absolutos de los elementos. Otra es la norma *magnitud-máxima* o *norma vector-uniforme*.

$$\|X\|_\infty = \max_{1 \le i \le n} |x_i|$$

la cual define la norma como el elemento con el mayor valor absoluto.

Utilizando un método similar, se pueden desarrollar normas para matrices. Por ejemplo,

$$\|A\|_1 = \max_{1 \le j \le n} \sum_{i=1}^{n} |a_{ij}|$$

Esto es, se realiza una sumatoria de los valores absolutos de los coeficientes para cada columna, y la mayor de estas sumatorias se toma como la norma. Esto se conoce como la *norma columna-suma*.

Una determinación similar se puede hacer para los renglones, y resulta una *matriz-uniforme* o *norma renglón-suma*,

$$\|A\|_\infty = \max_{1 \le i \le n} \sum_{j=1}^{n} |a_{ij}|$$

Debe observarse que, en contraste con los vectores, la norma 2 y la norma euclidiana para una matriz no son lo mismo. Mientras que la norma euclidiana $\|A\|_e$ puede ser fácilmente determinada mediante la ecuación (10.24), la norma 2 para matrices $\|A\|_2$ se calcula así:

$$\|A\|_2 = (\mu_{\text{máx}})^{1/2}$$

donde $\mu_{\text{máx}}$ es el mayor valor característico de $[A]^T[A]$. En el capítulo 27 se verá más sobre valores característicos. Mientras tanto, el punto importante es que la norma $\|A\|_2$, o *norma espectral*, es la norma mínima y, por lo tanto, proporciona la medida de tamaño más ajustada (Ortega, 1972).

$$\frac{\|\Delta X\|}{\|X\|} \leq \text{Cond } [A] \frac{\|\Delta A\|}{\|A\|}$$

Es decir, el error relativo de la norma de la solución calculada puede ser tan grande como el error relativo de la norma de los coeficientes de [A], multiplicada por el número de condición. Por ejemplo, si los coeficientes de [A] se encuentran a t dígitos de precisión (esto es, los errores de redondeo son del orden de 10^{-t}) y Cond [A] = 10^c, la solución [X] puede ser válida sólo para $t - c$ dígitos (errores de redondeo ~ 10^{c-t}).

EJEMPLO 10.4 Evaluación de la condición de una matriz

Planteamiento del problema La matriz de Hilbert, que es notoriamente mal condicionada, se representa como

$$\begin{bmatrix} 1 & 1/2 & 1/3 & \cdots & 1/n \\ 1/2 & 1/3 & 1/4 & \cdots & 1/(n+1) \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 1/n & 1/(n+1) & 1/(n+2) & \cdots & 1/(2n-1) \end{bmatrix}$$

Use la norma renglón-suma para estimar el número de condición de la matriz de Hilbert de 3×3,

$$[A] = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}$$

Solución Primero, la matriz se normaliza de tal forma que el elemento máximo en cada renglón sea 1.

$$[A] = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1 & 2/3 & 1/2 \\ 1 & 3/4 & 3/5 \end{bmatrix}$$

Sumando cada uno de los renglones el resultado es 1.833, 2.1667 y 2.35. Entonces, el tercer renglón tiene la suma mayor y la norma renglón-suma es

$$\|A\|_\infty = 1 + \frac{3}{4} + \frac{3}{5} = 2.35$$

La inversa de la matriz escalada se calcula como

$$[A]^{-1} = \begin{bmatrix} 9 & -18 & 10 \\ -36 & 96 & -60 \\ 30 & -90 & 60 \end{bmatrix}$$

Observe que los elementos de esta matriz son mayores que los de la matriz original. Esto también se refleja en su norma renglón-suma, la cual se calcula como

$$\|A^{-1}\|_\infty = |-36| + |96| + |-60| = 192$$

Entonces, el número de condición se calcula como

$$\text{Cond } [A] = 2.35(192) = 451.2$$

El hecho de que el número de condición sea considerablemente mayor que la unidad sugiere que el sistema está mal condicionado. La importancia del mal condicionamiento puede ser cuantificado

al calcular $c = \log 451.2 = 2.65$. Las computadoras que usan una representación de punto flotante IEEE tienen aproximadamente $t = \log 2^{-24} = 7.2$ dígitos significativos en base 10 (recuerde la sección 3.4.1). Por lo tanto, la solución puede tener errores de redondeo de hasta $10^{(2.65-7.2)} = 3 \times 10^{-5}$. Observe que una estimación como ésta casi siempre sobrepredice el error verdadero. Sin embargo, son útiles para alertar al usuario en el caso de que los errores de redondeo puedan resultar significativos.

En pocas palabras, el problema al usar la ecuación (10.26) es el precio computacional requerido para obtener $\|A^{-1}\|$. Rice (1983) indica algunas posibles estrategias para reducir el problema. Además, él sugiere una forma alternativa para determinar la condición del sistema: ejecute la misma solución en dos diferentes compiladores. Ya que los códigos resultantes implementan en forma diferente la aritmética, el efecto de mal condicionamiento debería ser evidente en un experimento como ése. Por último, se debe mencionar que los paquetes de software, como MATLAB y Mathcad, tienen la capacidad para calcular en forma conveniente la condición de una matriz. Revisaremos estas capacidades cuando se vean esos paquetes al final del capítulo 11.

10.3.3 Refinamiento iterativo

En algunos casos, los errores de redondeo se reducen con el siguiente procedimiento. Suponga que se está resolviendo el siguiente sistema de ecuaciones:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= b_3 \end{aligned} \tag{10.27}$$

Se limitará el siguiente análisis a un sistema pequeño de (3 × 3). Aunque este método se puede generalizar para aplicarlo a sistemas de ecuaciones lineales más grandes.

Suponga que una solución aproximada en forma vectorial es $\{\tilde{X}\}^T = \lfloor \tilde{x}_1 \;\; \tilde{x}_2 \;\; \tilde{x}_3 \rfloor$. Esta solución se sustituye en la ecuación (10.27) para tener

$$\begin{aligned} a_{11}\tilde{x}_1 + a_{12}\tilde{x}_2 + a_{13}\tilde{x}_3 &= \tilde{b}_1 \\ a_{21}\tilde{x}_1 + a_{22}\tilde{x}_2 + a_{23}\tilde{x}_3 &= \tilde{b}_2 \\ a_{31}\tilde{x}_1 + a_{32}\tilde{x}_2 + a_{33}\tilde{x}_3 &= \tilde{b}_3 \end{aligned} \tag{10.28}$$

Ahora, suponga que la solución exacta $\{X\}$ está expresada como una función de la solución aproximada y de un vector de factores de corrección $\{\Delta X\}$, donde

$$\begin{aligned} x_1 &= \tilde{x}_1 + \Delta x_1 \\ x_2 &= \tilde{x}_2 + \Delta x_2 \\ x_3 &= \tilde{x}_3 + \Delta x_3 \end{aligned} \tag{10.29}$$

Estos resultados se sustituyen en la ecuación (10.27), para obtener el siguiente sistema:

$$\begin{aligned} a_{11}(\tilde{x}_1 + \Delta x_1) + a_{12}(\tilde{x}_2 + \Delta x_2) + a_{13}(\tilde{x}_3 + \Delta x_3) &= b_1 \\ a_{21}(\tilde{x}_1 + \Delta x_1) + a_{22}(\tilde{x}_2 + \Delta x_2) + a_{23}(\tilde{x}_3 + \Delta x_3) &= b_2 \\ a_{31}(\tilde{x}_1 + \Delta x_1) + a_{32}(\tilde{x}_2 + \Delta x_2) + a_{33}(\tilde{x}_3 + \Delta x_3) &= b_3 \end{aligned} \tag{10.30}$$

Ahora, la ecuación (10.28) se resta de la (10.30) para dar

$$\begin{aligned} a_{11}\Delta x_1 + a_{12}\Delta x_2 + a_{13}\Delta x_3 &= b_1 - \tilde{b}_1 = E_1 \\ a_{21}\Delta x_1 + a_{22}\Delta x_2 + a_{23}\Delta x_3 &= b_2 - \tilde{b}_2 = E_2 \\ a_{31}\Delta x_1 + a_{32}\Delta x_2 + a_{33}\Delta x_3 &= b_3 - \tilde{b}_3 = E_3 \end{aligned} \tag{10.31}$$

Así este sistema es un conjunto de ecuaciones lineales simultáneas que puede resolverse para obtener los factores de corrección. Dichos factores se aplican para mejorar la solución, como lo especifica la ecuación (10.29).

Es relativamente sencillo agregar un procedimiento de refinamiento iterativo en los programas de computadora para métodos de eliminación. Esto es especialmente efectivo para los métodos de descomposición *LU* descritos antes, los cuales sirven para evaluar en forma eficiente varios vectores del lado derecho. Observe que para ser efectivos en sistemas mal condicionados, las *E* en la ecuación (10.31) deben expresarse en doble precisión.

PROBLEMAS

10.1 Utilice las reglas de la multiplicación de matrices para demostrar que las ecuaciones (10.7) y (10.8) se obtienen de la (10.6).

10.2 *a*) Use la eliminación simple de Gauss para descomponer el sistema siguiente, de acuerdo con la descripción de la sección 10.1.2.

$$10x_1 + 2x_2 - x_3 = 27$$
$$-3x_1 - 6x_2 + 2x_3 = -61.5$$
$$x_1 + x_2 + 5x_3 = -21.5$$

Después, multiplique las matrices [*L*] y [*U*] resultantes para demostrar que se genera [*A*]. *b*) Emplee la descomposición *LU* para resolver el sistema. Realice todos los pasos del cálculo. *c*) También resuelva el sistema para un vector alternativo del lado derecho: $\{B\}^T = [12 \;\; 18 \;\; -6]$.

10.3

a) Resuelva el sistema de ecuaciones siguiente por medio de la descomposición *LU* sin pivoteo.

$$8x_1 + 4x_2 - x_3 = 11$$
$$-2x_1 + 5x_2 + x_3 = 4$$
$$2x_1 - x_2 + 6x_3 = 7$$

b) Determine la matriz inversa. Compruebe sus resultados por medio de verificar que $[A][A]^{-1} = [I]$.

10.4 Resuelva el sistema de ecuaciones siguiente por medio de la descomposición *LU* con pivoteo parcial:

$$2x_1 - 6x_2 - x_3 = -38$$
$$-3x_1 - x_2 + 7x_3 = -34$$
$$-8x_1 + x_2 - 2x_3 = -20$$

10.5 Determine los flops totales como función del número de ecuaciones *n* para las fases de *a*) descomposición, *b*) sustitución hacia adelante y *c*) sustitución hacia atrás, de la versión de la descomposición *LU* de la eliminación de Gauss.

10.6 Utilice la descomposición *LU* para determinar la matriz inversa del sistema que sigue. No use una estrategia de pivoteo, y compruebe su resultado con la verificación de que $[A][A]^{-1} = [I]$.

$$10x_1 + 2x_2 - x_3 = 27$$
$$-3x_1 - 6x_2 - 2x_3 = -61.5$$
$$x_1 + x_2 + 5x_3 = -21.5$$

10.7 Ejecute la descomposición de Crout sobre el sistema

$$2x_1 - 5x_2 + x_3 = 12$$
$$-x_1 + 3x_2 - x_3 = -8$$
$$3x_1 - 4x_2 + 2x_3 = 16$$

Después, multiplique las matrices [*L*] y [*U*] resultantes para determinar que se produce [*A*].

10.8 El sistema de ecuaciones que sigue está diseñado para determinar concentraciones (las *c* están en g/m³) en una serie de reactores acoplados, como función de la cantidad de masa que entra a cada uno de ellos (los lados derechos están en g/día),

$$15c_1 - 3c_2 - c_3 = 3\,800$$
$$-3c_1 + 18c_2 - 6c_3 = 1\,200$$
$$-4c_1 - c_2 + 12c_3 = 2\,350$$

a) Determine la matriz inversa.

b) Use la inversa para encontrar la solución.

c) Determine cuánto debe incrementarse la tasa de masa de entrada al reactor 3 para inducir un aumento de 10 g/m³ en la concentración del reactor 1.

d) ¿Cuánto se reduciría la concentración en el reactor 3 si la tasa de masa de entrada a los reactores 1 y 2 se redujera en 500 y 250 g/día, respectivamente?

10.9 Resuelva el siguiente sistema de ecuaciones con descomposición *LU*:

$$3x_1 - 2x_2 + x_3 = -10$$
$$2x_1 + 6x_2 - 4x_3 = 44$$
$$-x_1 - 2x_2 + 5x_3 = -26$$

10.10 *a*) Determine la descomposición *LU* para la siguiente matriz, sin usar una estrategia de pivoteo, y verifique sus resultados validando que [*L*][*U*] = [*A*].

$$\begin{bmatrix} 8 & 2 & 1 \\ 3 & 7 & 2 \\ 2 & 3 & 9 \end{bmatrix}$$

b) Utilice el resultado de *a*) para calcular el determinante.

c) Repita *a*) y *b*) usando MATLAB.

10.11 Use la siguiente descomposición *LU* para *a*) calcular el determinante y *b*) resolver $[A]\{x\} = \{b\}$ con $\{b\}^T = [-10 \;\; 44 \;\; -26]$.

$$[A] = [L][U] = \begin{bmatrix} 1 & & \\ 0.6667 & 1 & \\ -0.3333 & -0.3636 & 1 \end{bmatrix} \begin{bmatrix} 3 & -2 & 1 \\ & 7.3333 & -4.6667 \\ & & 3.6364 \end{bmatrix}$$

10.12 Determine $\|A\|_e$, $\|A\|_1$ y $\|A\|_\infty$ para

$$[A] = \begin{bmatrix} 8 & 2 & -10 \\ -9 & 1 & 3 \\ 15 & -1 & 6 \end{bmatrix}$$

Escale la matriz haciendo que el máximo elemento de cada renglón sea igual a uno.

10.13 Determine las normas de Frobenius y de renglón-suma para los sistemas de los problemas 10.3 y 10.4. Escale las matrices igualando a uno el elemento máximo de cada renglón.

10.14 Una matriz [*A*] está definida como sigue

$$[A] = \begin{bmatrix} 0.125 & 0.25 & 0.5 & 1 \\ 0.015625 & 0.625 & 0.25 & 1 \\ 0.00463 & 0.02777 & 0.16667 & 1 \\ 0.001953 & 0.015625 & 0.125 & 1 \end{bmatrix}$$

Con el uso de la norma renglón-suma, calcule el número de condición y cuántos dígitos sospechosos se generarían con esta matriz.

10.15 *a*) Determine el número de condición para el sistema siguiente por medio de la norma renglón-suma. No normalice el sistema.

$$\begin{bmatrix} 1 & 4 & 9 & 16 & 25 \\ 4 & 9 & 16 & 25 & 36 \\ 9 & 16 & 25 & 36 & 49 \\ 16 & 25 & 36 & 49 & 64 \\ 25 & 36 & 49 & 64 & 81 \end{bmatrix}$$

¿Cuántos dígitos de precisión se perderían debido al mal acondicionamiento? *b*) Repita el inciso *a*), pero escale la matriz por medio de hacer el elemento más grande de cada renglón igual a uno.

10.16 Determine el número de condición con base en la norma renglón-suma para la matriz de Hilbert normalizada de 5×5. ¿Cuántos dígitos significativos de precisión se perderían debido al mal acondicionamiento?

10.17 Además de la matriz de Hilbert, hay otras matrices que son inherentemente mal acondicionadas. Uno de esos casos es la *matriz de Vandermonde*, que tiene la forma siguiente:

$$\begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ x_3^2 & x_3 & 1 \end{bmatrix}$$

a) Determine el número de condición con base en la norma renglón-suma para el caso en que $x_1 = 4$, $x_2 = 2$ y $x_3 = 7$.
b) Emplee el software de MATLAB o Mathcad para calcular los números de condición espectral y de Frobenius.

10.18 Desarrolle un programa amigable para el usuario para hacer la descomposición *LU* con base en el pseudocódigo de la figura 10.2.

10.19 Realice un programa amigable para el usuario para efectuar la descomposición *LU*, que incluya la capacidad de evaluar la matriz inversa. Fundamente el programa en las figuras 10.2 y 10.5.

10.20 Use técnicas iterativas de refinamiento para mejorar $x_1 = 2$, $x_2 = -3$ y $x_3 = 8$, que son las soluciones aproximadas de

$$2x_1 + 5x_2 + x_3 = -5$$
$$5x_1 + 2x_2 + x_3 = 12$$
$$x_1 + 2x_2 + x_3 = 3$$

10.21 Considere los vectores siguientes:

$$\vec{A} = 2\vec{\boldsymbol{i}} - 3\vec{\boldsymbol{j}} + a\vec{\boldsymbol{k}}$$
$$\vec{B} = b\vec{\boldsymbol{i}} + \vec{\boldsymbol{j}} - 4\vec{\boldsymbol{k}}$$
$$\vec{C} = 3\vec{\boldsymbol{i}} + c\vec{\boldsymbol{j}} + 2\vec{\boldsymbol{k}}$$

El vector $\vec{A}$ es perpendicular al $\vec{B}$ y al $\vec{C}$. También se sabe que $\vec{B}\cdot\vec{C} = 2$. Use cualquier método de los estudiados en este capítulo para resolver las tres incógnitas, *a*, *b* y *c*.

10.22 Considere los vectores siguientes:

$$\vec{A} = a\vec{\boldsymbol{i}} + b\vec{\boldsymbol{j}} + c\vec{\boldsymbol{k}}$$
$$\vec{B} = -2\vec{\boldsymbol{i}} + \vec{\boldsymbol{j}} - 4\vec{\boldsymbol{k}}$$
$$\vec{C} = \vec{\boldsymbol{i}} + 3\vec{\boldsymbol{j}} + 2\vec{\boldsymbol{k}}$$

donde $\vec{A}$ es un vector desconocido. Si

$$(\vec{A}\times\vec{B}) + (\vec{A}\times\vec{C}) = (5a+6)\vec{\boldsymbol{i}} + (3b-2)\vec{\boldsymbol{j}} + (-4c+1)\vec{\boldsymbol{k}}$$

use cualquier método de los que aprendió en este capítulo para resolver para las tres incógnitas, *a*, *b* y *c*.

10.23 Deje que la función esté definida en el intervalo [0, 2] como sigue:

$$f(x) = \begin{cases} ax + b, & 0 \le x \le 1 \\ cx + d, & 1 \le x \le 2 \end{cases}$$

Determine las constantes *a*, *b*, *c* y *d*, de modo que la función *f* satisfaga lo siguiente:

- $f(0) = f(2) = 1$.
- *f* es continua en todo el intervalo.
- $a + b = 4$.

Obtenga y resuelva un sistema de ecuaciones algebraicas lineales con una forma matricial idéntica a la ecuación (10.1).

10.24 *a*) Cree una matriz de Hilbert de 3×3. Ésta será la matriz [*A*]. Multiplique la matriz por el vector columna $\{x\} = [1, 1, 1]^T$. La solución de $[A]\{x\}$ será otro vector columna $\{b\}$. Con el uso de cualquier paquete numérico y la eliminación de Gauss, encuentre la solución de $[A]\{x\} = \{b\}$ por medio del empleo de la matriz de Hilbert y el vector $\{b\}$ que calculó. Compare el resultado con su vector $\{x\}$ conocido. Utilice precisión suficiente al mostrar los resultados con objeto de permitir detectar imprecisiones.
b) Repita el inciso *a*) con el uso de una matriz de Hilbert de 7×7.
c) Repita el inciso *a*) con el uso de una matriz de Hilbert de 10×10.

10.25 La interpolación polinómica consiste en determinar el único polinomio de grado ($n - 1$) que se acomode a *n* puntos de datos. Tales funciones polinómicas tienen la forma general

$$f(x) = p_1x^{n-1} + p_2x^{n-2} + \cdots + p_{n-1}x + p_n \qquad \text{(P10.25)}$$

donde las *p* son coeficientes constantes. Una forma directa de calcular los coeficientes es generar *n* ecuaciones algebraicas lineales que se puedan resolver simultáneamente para despejar los coeficientes. Suponga que se quiere determinar los coeficientes de la función polinómica de cuarto grado $f(x) = p_1x^4 + p_2x^3 + p_3x^2 + p_4x + p_5$ que pasa por los siguientes cinco puntos: (200, 0.746), (250, 0.675), (300, 0.616), (400, 0.525) y (500, 0.457). Cada uno de estos pares se puede sustituir en la ecuación (P10.25) para dar un sistema de cinco ecuaciones con cinco incógnitas (las *p*). Use este procedimiento para despejar los coeficientes. Además, determine e interprete el número de condición.

CAPÍTULO 11

Matrices especiales y el método de Gauss-Seidel

Ciertas matrices tienen una estructura particular que puede aprovecharse para desarrollar esquemas de solución eficientes. La primera parte de este capítulo se dedica a dos de estos sistemas: *matrices a bandas* y *simétricas*. Se describen métodos de eliminación eficiente para ambas.

La segunda parte de este capítulo presenta una alternativa a los métodos de eliminación, es decir, métodos iterativos. El enfoque se da con el *método de Gauss-Seidel*, el cual emplea valores iniciales y después itera para obtener mejores aproximaciones a la solución. El método de Gauss-Seidel es particularmente adecuado cuando se tiene gran número de ecuaciones. En estos casos, los métodos de eliminación pueden estar sujetos a errores de redondeo. Debido a que el error en el método de Gauss-Seidel es determinado por el número de iteraciones, el error de redondeo no es un tema que preocupe a este método. Sin embargo, existen ciertos ejemplos donde la técnica de Gauss-Seidel no convergerá al resultado correcto. Éstas y algunas otras ventajas y desventajas que se tienen entre los métodos de eliminación e iterativos se analizarán en las páginas siguientes.

11.1 MATRICES ESPECIALES

Como se mencionó en el cuadro PT3.1, una *matriz a bandas* es una matriz cuadrada en la que todos sus elementos son cero, con excepción de una banda centrada sobre la diagonal principal. Los sistemas a bandas se encuentran con frecuencia en la práctica científica y de la ingeniería. Por ejemplo, tales sistemas aparecen en la solución de ecuaciones diferenciales. Además, otros métodos numéricos como el de los splines cúbicos (sección 18.5) involucran la solución de sistemas a bandas.

Las dimensiones de un sistema a bandas se cuantifica mediante dos parámetros: el ancho de banda (BW, por sus iniciales en inglés) y el ancho de media banda HBW (figura 11.1). Estos dos valores se relacionan mediante BW = 2HBW + 1. En general, un sistema a bandas es aquel para el cual $a_{ij} = 0$ si $|i - j| > \text{HBW}$.

Aunque la eliminación de Gauss o la descomposición *LU* convencional se emplean para resolver sistemas de ecuaciones bandeados, resultan ser ineficientes, debido a que si el pivoteo no es necesario, ninguno de los elementos fuera de la banda cambiará su valor original igual a cero. Así, no será necesario utilizar tiempo y espacio en el almacenamiento y en el manejo de estos ceros inútiles. Si se sabe de antemano que el pivoteo no es necesario, se pueden desarrollar algoritmos muy eficientes en los que no intervengan los ceros fuera de la banda. Como en muchos problemas con sistemas a bandas, no se requiere el pivoteo; los algoritmos alternos, que se describirán a continuación, son los métodos seleccionados para tal fin.

11.1.1 Sistemas tridiagonales

Un sistema tridiagonal (es decir, uno con un ancho de banda 3) se expresa en forma general de la siguiente manera:

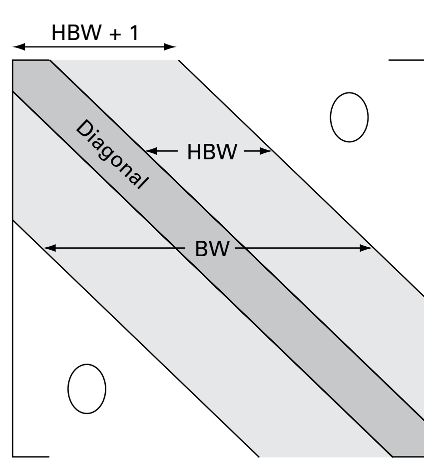


Figura 11.1
Parámetros utilizados para cuantificar las dimensiones de un sistema a bandas. BW y HBW designan el ancho de banda y el ancho de media banda, respectivamente.

a) **Descomposición**

```
DOFOR k = 2, n
   e_k = e_k/f_k-1
   f_k = f_k - e_k · g_k-1
END DO
```

b) **Sustitución hacia adelante**

```
DOFOR k = 2, n
   r_k = r_k - e_k · r_k-1
END DO
```

c) **Sustitución hacia atrás**

```
x_n = r_n/f_n
DOFOR k = n - 1, 1, -1
   x_k = (r_k - g_k · x_k+1)/f_k
END DO
```

Figura 11.2
Pseudocódigo para implementar el algoritmo de Thomas, un método de descomposición *LU* para sistemas tridiagonales.

$$\begin{bmatrix} f_1 & g_1 & & & & & \\ e_2 & f_2 & g_2 & & & & \\ & e_3 & f_3 & g_3 & & & \\ & & \cdot & \cdot & \cdot & & \\ & & & \cdot & \cdot & \cdot & \\ & & & & \cdot & \cdot & \cdot \\ & & & & e_{n-1} & f_{n-1} & g_{n-1} \\ & & & & & e_n & f_n \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ \cdot \\ \cdot \\ x_{n-1} \\ x_n \end{Bmatrix} = \begin{Bmatrix} r_1 \\ r_2 \\ r_3 \\ \cdot \\ \cdot \\ \cdot \\ r_{n-1} \\ r_n \end{Bmatrix} \tag{11.1}$$

Observe que se ha cambiado la notación para los coeficientes; en lugar de *a* y *b* usamos *e*, *f*, *g* y *r*. Esto se hace para evitar guardar un gran número de ceros que no se utilizan en la matriz cuadrada de las *a*. Esta modificación es ventajosa para ahorrar espacio, ya que el algoritmo resultante requiere menos memoria de cómputo.

En la figura 11.2 se muestra el pseudocódigo de un método eficiente, llamado *algoritmo de Thomas*, para resolver la ecuación (11.1). Como una descomposición *LU* convencional, el algoritmo consiste de tres pasos: descomposición, sustitución hacia adelante y sustitución hacia atrás. Así, las ventajas de la descomposición *LU*, como la evaluación de vectores múltiples del lado derecho y el cálculo de la matriz inversa, se obtienen mediante una apropiada aplicación de este algoritmo.

EJEMPLO 11.1 Solución tridiagonal con el algoritmo de Thomas

Planteamiento del problema Resuelva el siguiente sistema tridiagonal con el algoritmo de Thomas.

$$\begin{bmatrix} 2.04 & -1 & & \\ -1 & 2.04 & -1 & \\ & -1 & 2.04 & -1 \\ & & -1 & 2.04 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{Bmatrix} = \begin{Bmatrix} 40.8 \\ 0.8 \\ 0.8 \\ 200.8 \end{Bmatrix}$$

Solución Primero, la descomposición se realiza así:

$$e_2 = -1/2.04 = -0.49$$

$$f_2 = 2.04 - (-0.49)(-1) = 1.550$$

$$e_3 = -1/1.550 = -0.645$$

$$f_3 = 2.04 - (-0.645)(-1) = 1.395$$

$$e_4 = -1/1.395 = -0.717$$

$$f_4 = 2.04 - (-0.717)(-1) = 1.323$$

Así, la matriz se transforma en

$$\begin{bmatrix} 2.04 & -1 & & \\ -0.49 & 1.550 & -1 & \\ & -0.645 & 1.395 & -1 \\ & & -0.717 & 1.323 \end{bmatrix}$$

y la descomposición *LU* es

$$[A] = [L][U] = \begin{bmatrix} 1 & & & \\ -0.49 & 1 & & \\ & -0.645 & 1 & \\ & & -0.717 & 1 \end{bmatrix} \begin{bmatrix} 2.04 & -1 & & \\ & 1.550 & -1 & \\ & & 1.395 & -1 \\ & & & 1.323 \end{bmatrix}$$

Se puede verificar que ésta sea correcta al multiplicar $[L][U]$ para obtener $[A]$.

La sustitución hacia adelante se realiza de la siguiente manera:

$r_2 = 0.8 - (-0.49)40.8 = 20.8$

$r_3 = 0.8 - (-0.645)20.8 = 14.221$

$r_4 = 200.8 - (-0.717)14.221 = 210.996$

De esta forma, el vector del lado derecho se modifica a

$$\begin{Bmatrix} 40.8 \\ 20.8 \\ 14.221 \\ 210.996 \end{Bmatrix}$$

el cual, entonces, se utiliza de manera conjunta con la matriz $[U]$, para realizar la sustitución hacia atrás y obtener la solución

$T_4 = 210.996/1.323 = 159.480$

$T_3 = [14.221 - (-1)159.48]/1.395 = 124.538$

$T_2 = [20.800 - (-1)124.538]/1.550 = 93.778$

$T_1 = [40.800 - (-1)93.778]/2.040 = 65.970$

11.1.2 Descomposición de Cholesky

Recuerde del cuadro PT3.1, que una matriz simétrica es aquella donde $a_{ij} = a_{ji}$ para toda i y j. En otras palabras, $[A] = [A]^T$. Tales sistemas se presentan comúnmente en problemas de contexto matemático y de ingeniería. Estas matrices ofrecen ventajas computacionales, ya que únicamente se necesita la mitad de espacio de almacenamiento y, en la mayoría de los casos, sólo se requiere la mitad del tiempo de cálculo para su solución.

Uno de los métodos más populares usa la *descomposición de Cholesky*. Este algoritmo se basa en el hecho de que una matriz simétrica se descompone así:

$$[A] = [L][L]^T \tag{11.2}$$

Es decir, los factores triangulares resultantes son la transpuesta uno de otro.

Los términos de la ecuación (11.2) se desarrollan al multiplicar e igualar entre sí ambos lados. El resultado se expresa en forma simple mediante relaciones de recurrencia. Para el renglón k-ésimo,

$$l_{ki} = \frac{a_{ki} - \sum_{j=1}^{i-1} l_{ij} l_{kj}}{l_{ii}} \quad \text{para } i = 1, 2, \ldots, k-1 \tag{11.3}$$

y

$$l_{kk} = \sqrt{a_{kk} - \sum_{j=1}^{k-1} l_{kj}^2} \tag{11.4}$$

EJEMPLO 11.2 Descomposición de Cholesky

Planteamiento del problema Aplique la descomposición de Cholesky a la matriz simétrica

$$[A]=\begin{bmatrix}6 & 15 & 55\\ 15 & 55 & 225\\ 55 & 225 & 979\end{bmatrix}$$

Solución Para el primer renglón ($k = 1$), no se toma en cuenta la ecuación (11.3) y se emplea la ecuación (11.4) para calcular

$$l_{11}=\sqrt{a_{11}}=\sqrt{6}=2.4495$$

Para el segundo renglón ($k = 2$), con la ecuación (11.3) se obtiene

$$l_{21}=\frac{a_{21}}{l_{11}}=\frac{15}{2.4495}=6.1237$$

y con la ecuación (11.4) resulta

$$l_{22}=\sqrt{a_{22}-l_{21}^2}=\sqrt{55-(6.1237)^2}=4.1833$$

Para el tercer renglón ($k = 3$), la ecuación (11.3) con $i = 1$ da como resultado

$$l_{31}=\frac{a_{31}}{l_{11}}=\frac{55}{2.4495}=22.454$$

y con ($i = 2$)

$$l_{32}=\frac{a_{32}-l_{21}l_{31}}{l_{22}}=\frac{225-6.1237(22.454)}{4.1833}=20.917$$

en la ecuación (11.4) se obtiene

$$l_{33}=\sqrt{a_{33}-l_{31}^2-l_{32}^2}=\sqrt{979-(22.454)^2-(20.917)^2}=6.1101$$

De esta forma, la descomposición de Cholesky queda como

$$[L]=\begin{bmatrix}2.4495 & & \\ 6.1237 & 4.1833 & \\ 22.454 & 20.917 & 6.1101\end{bmatrix}$$

Se verifica la validez de esta descomposición al sustituirla junto con su transpuesta en la ecuación (11.2) y ver que del producto resulta la matriz original $[A]$. Esto se deja como ejercicio para el lector.

En la figura 11.3 se presenta el pseudocódigo para el algoritmo de la descomposición de Cholesky. Debe observar que el algoritmo de la figura 11.3 da un error de ejecución si en la evaluación de a_{kk} se obtiene la raíz cuadrada de un número negativo. Sin embargo, cuando la matriz es *definida positiva*,[1]

[1] Una *matriz definida* positiva es aquella para la cual el producto $\{X\}^T[A]\{X\}$ es mayor que cero, para todo vector $\{X\}$ distinto de cero.

esto nunca ocurrirá. Debido a que muchas de las matrices simétricas que se usan en ingeniería son de hecho definidas positivas, el algoritmo de Cholesky tiene una amplia aplicación. Otro beneficio al trabajar con matrices simétricas definidas positivas es que no se requiere el pivoteo para evitar la división entre cero. Así, es posible implementar el algoritmo de la figura 11.3 sin la complicación del pivoteo.

```
DOFOR k = 1, n
 DOFOR i = 1, k - 1
  sum = 0.
  DOFOR j = 1, i - 1
   sum = sum + a_ij · a_kj
  END DO
  a_ki = (a_ki - sum)/a_ii
 END DO
 sum = 0.
 DOFOR j = 1, k - 1
  sum = sum + a^2_kj
 END DO
 a_kk = √(a_kk - sum)
END DO
```

Figura 11.3
Pseudocódigo para el algoritmo de descomposición *LU* de Cholesky.

11.2 GAUSS-SEIDEL

Los métodos iterativos constituyen una alternativa a los métodos de eliminación descritos hasta ahora, para aproximar la solución. Tales métodos son similares a las técnicas que se desarrollaron en el capítulo 6 para obtener las raíces de una sola ecuación. Aquellos métodos consistían en suponer un valor y luego usar un método sistemático para obtener una aproximación mejorada de la raíz. Como esta parte del libro trata con un problema similar (obtener los valores que simultáneamente satisfagan un conjunto de ecuaciones). Entonces se esperaría que tales métodos aproximados fuesen útiles en este contexto.

El *método de Gauss-Seidel* es el método iterativo más comúnmente usado. Suponga que se da un sistema de n ecuaciones:

$$[A]\{X\} = \{B\}$$

Suponga que se limita a un conjunto de ecuaciones de 3 × 3. Si los elementos de la diagonal no son todos cero, la primera ecuación se puede resolver para x_1, la segunda para x_2 y la tercera para x_3, para obtener

$$x_1 = \frac{b_1 - a_{12}x_2 - a_{13}x_3}{a_{11}} \tag{11.5a}$$

$$x_2 = \frac{b_2 - a_{21}x_1 - a_{23}x_3}{a_{22}} \tag{11.5b}$$

$$x_3 = \frac{b_3 - a_{31}x_1 - a_{32}x_2}{a_{33}} \tag{11.5c}$$

Ahora, se puede empezar el proceso de solución al escoger valores iniciales para las x. Una forma simple para obtener los valores iniciales es suponer que todos son cero. Estos ceros se sustituyen en la ecuación (11.5*a*), la cual se utiliza para calcular un nuevo valor $x_1 = b_1/a_{11}$. Después, se sustituye este nuevo valor de x_1 junto con el valor previo cero de x_3 en la ecuación (11.5*b*) y se calcula el nuevo valor de x_2. Este proceso se repite con la ecuación (11.5*c*) para calcular un nuevo valor de x_3. Después se regresa a la primera ecuación y se repite todo el procedimiento hasta que la solución converja suficientemente cerca a los valores verdaderos. La convergencia se verifica usando el criterio [recuerde la ecuación (3.5)]

$$|\varepsilon_{a,i}| = \left|\frac{x_i^j - x_i^{j-1}}{x_i^j}\right| 100\% < \varepsilon_s \tag{11.6}$$

para todas las i, donde j y $j - 1$ son las iteraciones actuales y previas, respectivamente.

EJEMPLO 11.3 Método de Gauss-Seidel

Planteamiento del problema Use el método de Gauss-Seidel para obtener la solución del sistema usado en el ejemplo 10.2:

$$3x_1 - 0.1x_2 - 0.2x_3 = 7.85$$
$$0.1x_1 + 7x_2 - 0.3x_3 = -19.3$$
$$0.3x_1 - 0.2x_2 + 10x_3 = 71.4$$

Recuerde que la solución exacta es $x_1 = 3$, $x_2 = -2.5$ y $x_3 = 7$.

Solución Primero, despeje la incógnita sobre la diagonal para cada una de las ecuaciones.

$$x_1 = \frac{7.85 + 0.1x_2 + 0.2x_3}{3} \tag{E11.3.1}$$

$$x_2 = \frac{-19.3 - 0.1x_1 + 0.3x_3}{7} \tag{E11.3.2}$$

$$x_3 = \frac{71.4 - 0.3x_1 + 0.2x_2}{10} \tag{E11.3.3}$$

Suponiendo que x_2 y x_3 son cero, se utiliza la ecuación (E11.3.1) para calcular

$$x_1 = \frac{7.85 + 0 + 0}{3} = 2.616667$$

Este valor, junto con el valor de $x_3 = 0$, se sustituye en la ecuación (E11.3.2) para calcular

$$x_2 = \frac{-19.3 - 0.1(2.616667) + 0}{7} = -2.794524$$

La primera iteración termina al sustituir los valores calculados para x_1 y x_2 en la ecuación (E11.3.3) para dar

$$x_3 = \frac{71.4 - 0.3(2.616667) + 0.2(-2.794524)}{10} = 7.005610$$

En la segunda iteración, se repite el mismo proceso para calcular

$$x_1 = \frac{7.85 + 0.1(-2.794524) + 0.2(7.005610)}{3} = 2.990557 \qquad |\varepsilon_t| = 0.31\%$$

$$x_2 = \frac{-19.3 - 0.1(2.990557) + 0.3(7.005610)}{7} = -2.499625 \qquad |\varepsilon_t| = 0.015\%$$

$$x_3 = \frac{71.4 - 0.3(2.990557) + 0.2(-2.499625)}{10} = 7.000291 \qquad |\varepsilon_t| = 0.0042\%$$

El método es, por lo tanto, convergente hacia la verdadera solución. Es posible aplicar iteraciones adicionales para mejorar los resultados. Sin embargo, en un problema real, no se podría saber *a priori* el resultado correcto. En consecuencia, la ecuación (11.6) nos da un medio para estimar el error. Por ejemplo, para x_1,

$$|\varepsilon_{a,1}| = \left| \frac{2.990557 - 2.616667}{2.990557} \right| 100\% = 12.5\%$$

Para x_2 y x_3, los errores estimados son $|\varepsilon_{a,2}| = 11.8\%$ y $|\varepsilon_{a,3}| = 0.076\%$. Observe que, como cuando se determinaron las raíces de una sola ecuación, las formulaciones como la ecuación (11.6) usualmente ofrecen una valoración conservativa de la convergencia. Así, cuando éstas se satisfacen, aseguran que el resultado se conozca con, al menos, la tolerancia especificada por ε_s.

Conforme un nuevo valor de x se calcula con el método de Gauss-Seidel, éste se usa inmediatamente en la siguiente ecuación para determinar el otro valor de x. De esta forma, si la solución es convergente, se empleará la mejor aproximación disponible. Un método alternativo, llamado *iteración de Jacobi*, emplea una táctica algo diferente. Más que usar inmediatamente el último valor disponible de x, esta técnica usa la ecuación (11.5) para calcular un conjunto de nuevas x con base en un conjunto de x anteriores. De esta forma, conforme se generan nuevos valores, no se usan en forma inmediata sino que se retienen para la siguiente iteración.

La diferencia entre el método de Gauss-Seidel y la iteración de Jacobi se muestra en la figura 11.4. Aunque hay ciertos casos donde es útil el método de Jacobi, la utilización de Gauss-Seidel da la mejor aproximación y usualmente lo hace el método preferido.

Primera iteración

a)	b)
$x_1 = (b_1 - a_{12}x_2 - a_{13}x_3)/a_{11}$	$x_1 = (b_1 - a_{12}x_2 - a_{13}x_3)/a_{11}$
$x_2 = (b_2 - a_{21}x_1 - a_{23}x_3)/a_{22}$	$x_2 = (b_2 - a_{21}x_1 - a_{23}x_3)/a_{22}$
$x_3 = (b_3 - a_{31}x_1 - a_{32}x_2)/a_{33}$	$x_3 = (b_3 - a_{31}x_1 - a_{32}x_2)/a_{33}$

Segunda iteración

a)	b)
$x_1 = (b_1 - a_{12}x_2 - a_{13}x_3)/a_{11}$	$x_1 = (b_1 - a_{12}x_2 - a_{13}x_3)/a_{11}$
$x_2 = (b_2 - a_{21}x_1 - a_{23}x_3)/a_{22}$	$x_2 = (b_2 - a_{21}x_1 - a_{23}x_3)/a_{22}$
$x_3 = (b_3 - a_{31}x_1 - a_{32}x_2)/a_{33}$	$x_3 = (b_3 - a_{31}x_1 - a_{32}x_2)/a_{33}$

Figura 11.4
Ilustración de la diferencia entre los métodos iterativos *a*) Gauss-Seidel y *b*) Jacobi para resolver ecuaciones algebraicas simultáneas lineales.

11.2.1 Criterio de convergencia para el método de Gauss-Seidel

Observe que el método de Gauss-Seidel es similar en esencia a la técnica de iteración de punto fijo que se usó en la sección 6.1 para obtener las raíces de una sola ecuación. Recuerde que la iteración de punto fijo presenta dos problemas fundamentales: 1) en algunas ocasiones no es convergente, y 2) cuando converge, con frecuencia lo hace en forma muy lenta. El método de Gauss-Seidel puede también presentar estas desventajas.

El criterio de convergencia se puede desarrollar al recordar de la sección 6.5.1 que las condiciones suficientes para la convergencia de dos ecuaciones no lineales, $u(x, y)$ y $v(x, y)$, son

$$\left|\frac{\partial u}{\partial x}\right| + \left|\frac{\partial u}{\partial y}\right| < 1 \tag{11.7a}$$

y

$$\left|\frac{\partial v}{\partial x}\right| + \left|\frac{\partial v}{\partial y}\right| < 1 \tag{11.7b}$$

Este criterio se aplica también a las ecuaciones lineales que se resuelven con el método de Gauss-Seidel. Por ejemplo, en el caso de dos ecuaciones simultáneas, el algoritmo de Gauss-Seidel [ecuación (11.5)] se expresa como

$$u(x_1, x_2) = \frac{b_1}{a_{11}} - \frac{a_{12}}{a_{11}} x_2 \tag{11.8a}$$

y

$$v(x_1, x_2) = \frac{b_2}{a_{22}} - \frac{a_{21}}{a_{22}} x_1 \tag{11.8b}$$

Las derivadas parciales de estas ecuaciones se evalúan con respecto a cada una de las incógnitas así:

$$\frac{\partial u}{\partial x_1} = 0 \qquad \frac{\partial v}{\partial x_2} = -\frac{a_{12}}{a_{11}}$$

y

$$\frac{\partial v}{\partial x_1} = -\frac{a_{21}}{a_{22}} \qquad \frac{\partial v}{\partial x_2} = 0$$

que se sustituyen en la ecuación (11.7) para dar

$$\left|\frac{a_{12}}{a_{11}}\right| < 1 \tag{11.9a}$$

y

$$\left|\frac{a_{21}}{a_{22}}\right| < 1 \tag{11.9b}$$

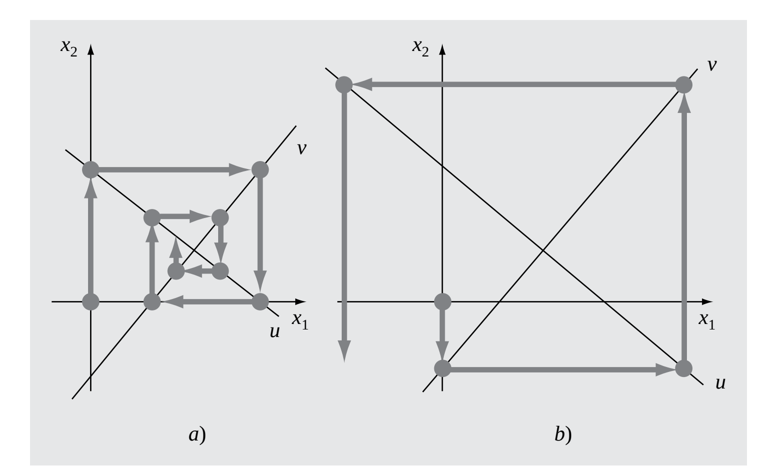


Figura 11.5
Representación *a*) de la convergencia y *b*) de la divergencia del método de Gauss-Seidel. Observe que las mismas funciones son dibujadas en ambos casos (*u*: $11x_1 + 13x_2 = 286$; *v*: $11x_1 - 9x_2 = 99$). Así, el orden en el cual se implementan las ecuaciones (se representa por la dirección de la primera flecha desde el origen) determina si el cálculo converge.

En otras palabras, el valor absoluto de las pendientes en la ecuación (11.8) debe ser menor que la unidad para asegurar la convergencia, lo cual se muestra gráficamente en la figura 11.5. La ecuación (11.9) también se reformula así:

$$|a_{11}| > |a_{12}|$$

y

$$|a_{22}| > |a_{21}|$$

Esto es, el elemento diagonal debe ser mayor que el elemento fuera de la diagonal para cada renglón.

La generalización de lo anterior para *n* ecuaciones es directa y se expresa como

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{i,j}| \tag{11.10}$$

Es decir, el coeficiente diagonal en cada una de las ecuaciones debe ser mayor que la suma del valor absoluto de los otros coeficientes de la ecuación. Este criterio es suficiente pero no necesario para la convergencia. Es decir, el método puede funcionar aunque no se satisfaga la ecuación (11.10), la convergencia se garantiza cuando la condición se satisface. A los sistemas que cumplen con la ecuación (11.10) se les conoce como *diagonalmente dominantes*. Por fortuna, en la ingeniería, muchos problemas de importancia práctica satisfacen este requerimiento.

11.2.2 Mejoramiento de la convergencia usando relajación

La *relajación* representa una ligera modificación al método de Gauss-Seidel y ésta permite mejorar la convergencia. Después de que se calcula cada nuevo valor de *x* por medio de la ecuación (11.5), ese valor se modifica mediante un promedio ponderado de los resultados de las iteraciones anterior y actual:

$$x_i^{\text{nuevo}} = \lambda x_i^{\text{nuevo}} + (1 - \lambda) x_i^{\text{anterior}} \tag{11.11}$$

donde λ es un factor ponderado que tiene un valor entre 0 y 2.

Si $\lambda = 1$, $(1 - \lambda)$ es igual a 0 y el resultado no se modifica. Sin embargo, si a λ se le asigna un valor entre 0 y 1, el resultado es un promedio ponderado de los resultados actuales y anteriores. Este tipo de modificación se conoce como *subrelajación*. Se emplea comúnmente para hacer que un sistema no convergente, converja o acelere la convergencia al amortiguar sus oscilaciones.

Para valores λ de 1 a 2, se le da una ponderación extra al valor actual. En este caso, hay una suposición implícita de que el nuevo valor se mueve en la dirección correcta hacia la solución verdadera, pero con una velocidad demasiado lenta. Por lo tanto, se pretende que la ponderación adicional de λ mejore la aproximación al llevarla más cerca de la verdadera. De aquí que este tipo de modificación, al cual se le llama *sobrerrelajación*, permite acelerar la convergencia de un sistema que ya es convergente. El método también se conoce como *sobrerrelajación simultánea* o *sucesiva*, o *SOR*.

La elección de un valor adecuado de λ es especificado por el problema y se determina en forma empírica. Para la solución de un solo sistema de ecuaciones, con frecuencia es innecesaria. No obstante, si el sistema bajo estudio se va a resolver de manera repetida, la eficiencia que se introduce por una prudente elección de λ puede ser extremadamente importante. Buenos ejemplos de lo anterior son los sistemas muy grandes de ecuaciones diferenciales parciales, que frecuentemente se presentan cuando se modelan variaciones continuas de variables (recuerde el sistema distribuido que se mostró en la figura PT3.1*b*). Se retomará el tema en la parte ocho.

11.2.3 Algoritmo para el método de Gauss-Seidel

En la figura 11.6 se muestra un algoritmo para el método de Gauss-Seidel con relajación. Observe que este algoritmo no garantiza la convergencia si las ecuaciones no se introducen en una forma diagonalmente dominante.

El pseudocódigo tiene dos características que vale la pena mencionar. La primera es que existe un conjunto inicial de ciclos anidados para dividir cada ecuación por su elemento diagonal. Esto reduce el número total de operaciones en el algoritmo. En la segunda, observe que la verificación del error se designa por una variable llamada *sentinel*. Si en cualquiera de las ecuaciones se tiene un error aproximado mayor que el criterio de paro (e_s), entonces se permite continuar con las iteraciones. El uso de la variable de estado ayuda a evitar cálculos innecesarios de estimación de error una vez que las ecuaciones excedan el criterio.

```
SUBROUTINE Gseid (a,b,n,x,imax,es,lambda)
  DOFOR i = 1,n
    dummy = a(i,i)
    DOFOR j = 1,n
      a(i,j) = a(i,j)/dummy
    END DO
    b(i) = b(i)/dummy
  END DO
  DOFOR i = 1, n
    sum = b(i)
    DOFOR j = 1, n
      IF i≠j THEN sum = sum – a(i,j)*x(j)
    END DO
    x(i)=sum
  END DO
  iter=1
  DO
    sentinel=1
    DOFOR i = 1,n
      old = x(i)
      sum = b(i)
      DOFOR j = 1,n
        IF i≠j THEN sum = sum – a(i,j)*x(j)
      END DO
      x(i) = lambda*sum +(1.-lambda)*old
      IF sentinel = 1 AND x(i) ≠ 0. THEN
        ea = ABS((x(i) – anterior)/x(i))*100.
        IF ea > es THEN sentinel = 0
      END IF
    END DO
    iter = iter + 1
    IF sentinel = 1 OR (iter ≥ imax) EXIT
  END DO
END Gseid
```

Figura 11.6
Pseudocódigo para el método de Gauss-Seidel con relajación.

11.2.4 Contextos del problema en el método de Gauss-Seidel

Además de evitar el problema de redondeo, la técnica de Gauss-Seidel tiene muchas otras ventajas que la hacen particularmente atractiva en el contexto de ciertos problemas de ingeniería. Por ejemplo, cuando la matriz en cuestión es muy grande y esparcida (es decir, cuando la mayoría de los elementos son cero), los métodos de eliminación desperdician grandes cantidades de memoria de cómputo al guardar ceros.

Al inicio de este capítulo se vio cómo esta desventaja se puede evitar si la matriz de coeficientes es a bandas. Para sistemas que no tienen la forma de banda, generalmente no existe una forma simple para evitar los grandes requerimientos de memoria cuando se utilizan los métodos de eliminación. Como todas las computadoras tienen una cantidad de memoria finita, esta ineficiencia llega a poner una limitación al tamaño de los sistemas, para los cuales los métodos de eliminación resultan prácticos.

Aunque un algoritmo general como el de la figura 11.6 es propenso a la misma restricción, la estructura de las ecuaciones de Gauss-Seidel [ecuación (11.5)] permite que se desarrollen programas concisos para sistemas específicos. Como sólo se necesita incluir coeficientes que no sean cero en la ecuación (11.5), es posible lograr grandes ahorros en la memoria de la computadora. Aunque esto implica más inversión en el desarrollo de software, las ventajas a largo plazo son sustanciales cuando se tiene grandes sistemas, en los cuales se ejecutan muchas simulaciones. Tanto sistemas de variables localizadas como distribuidas pueden dar como resultado matrices grandes y muy esparcidas donde el método de Gauss-Seidel tiene utilidad.

SOFTWARE

11.3 ECUACIONES ALGEBRAICAS LINEALES CON PAQUETES DE SOFTWARE

Los paquetes de software tienen grandes capacidades para resolver sistemas de ecuaciones algebraicas lineales. Antes de describir dichas herramientas, se debe mencionar que los procedimientos descritos en el capítulo 7 para resolver sistemas de ecuaciones no lineales pueden aplicarse a sistemas lineales. Sin embargo, en esta sección enfocaremos nuestra atención hacia procedimientos que están expresamente diseñados para ecuaciones lineales.

11.3.1 Excel

Existen dos formas para resolver ecuaciones algebraicas lineales con Excel: 1) por medio de la herramienta Solver o 2) usando la inversión de matrices y las funciones de multiplicación.

Recuerde que una forma para determinar la solución de ecuaciones algebraicas lineales es

$$\{X\} = [A]^{-1}\{B\} \tag{11.12}$$

Excel tiene funciones predeterminadas para inversión y multiplicación de matrices que sirve para implementar esta fórmula.

EJEMPLO 11.4 Uso de Excel para resolver sistemas lineales

Planteamiento del problema Recuerde que en el capítulo 10 se presentó la matriz de Hilbert. El siguiente sistema se basa en esta matriz. Observe que está escalado, como se realizó antes en el ejemplo 10.3, de tal forma que el coeficiente máximo en cada renglón es la unidad.

$$\begin{bmatrix} 1 & 1/2 & 1/3 \\ 1 & 2/3 & 1/2 \\ 1 & 3/4 & 3/5 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 1.833333 \\ 2.166667 \\ 2.35 \end{Bmatrix}$$

La solución a este sistema es $\{X\}^T = [1\ 1\ 1]$. Use Excel para obtener esta solución.

Solución La hoja de cálculo para resolver este problema se muestra en la figura 11.7. Primero, la matriz $[A]$ y las constantes del lado derecho $\{B\}$ se introducen en las celdas de la hoja de cálculo. Después, se resalta un conjunto de celdas de dimensiones apropiadas (en este ejemplo 3×3), ya sea presionando y arrastrando el ratón o por medio de las teclas direccionales mientras se presiona la tecla *shift*. Como se muestra en la figura 11.7, se resalta el rango: B5. .D7.

Ahora, se introduce una fórmula que invoca a la función matriz inversa,

```
=minverse(B1..D3)
```

Observe que el argumento es el rango que contiene los elementos de $[A]$. Las teclas *Ctrl* y *Shift* se mantienen presionadas mientras se oprime la tecla *Enter*. La inversa resultante de $[A]$ se calculará con Excel para desplegarse en el rango B5. .D7 como se muestra en la figura 11.7.

Un procedimiento similar se utiliza para multiplicar la inversa por el vector del lado derecho. En este caso, el rango de F5. .F7 se resalta y se introduce la siguiente fórmula

```
=mmult(B5..D7, F1..F3)
```

donde el primer rango es la primera matriz que habrá de multiplicarse, $[A]^{-1}$, y el segundo rango es la segunda matriz a multiplicarse, $\{B\}$. De nuevo, al usar la combinación *Ctrl-Shift-Enter*, la solución $\{X\}$ se calculará por Excel y desplegada en el rango F5. .F7, como se muestra en la figura 11.7. Como puede verse, se obtiene la respuesta correcta.

	A	B	C	D	E	F
1		1	0.5	0.33333333		1.83333333333333
2	[A] =	1	0.66666667	0.5	{B}=	2.16666666666667
3		1	0.75	0.6		2.35000000000000
4						
5		**9**	**–18**	**10**		**0.99999999999992**
6	[A]–1 =	**–36**	**96**	**–60**	**{X}=**	**1.00000000000043**
7		**30**	**–90**	**60**		**0.99999999999960**

=MINVERSE(B1:D3) **=MMULT(B5:D7,F1:F3)**

Figura 11.7

Observe que en forma deliberada se reformatearon los resultados en el ejemplo 11.4 para mostrar 15 dígitos. Esto se hizo debido a que Excel usa doble precisión para guardar valores numéricos. De esta forma, se observa que el error de redondeo ocurre en los últimos dos dígitos. Esto implica un número de condición del orden de 100, el cual concuerda con el resultado de 451.2 que originalmente se calculó en el ejemplo 10.3. Excel no tiene la capacidad para calcular un número de condición. En la mayoría de los casos, debido a que Excel emplea números con doble precisión, esto no representa un problema. Sin embargo, en casos donde se sospeche que el sistema esté mal condicionado, la determinación del número de condición es útil. MATLAB y Mathcad tienen la capacidad de calcular esta cantidad.

11.3.2 MATLAB

Como su nombre lo indica, MATLAB (contracción de MATrix LABoratory) se diseñó para facilitar el manejo de matrices. Así, como es de esperarse, sus capacidades en esta área son excelentes. Algunas funciones de MATLAB relacionadas con las operaciones de matrices se presentan en una lista en la tabla 11.1. El ejemplo siguiente ilustra algunas de dichas capacidades.

EJEMPLO 11.5 Uso de MATLAB para manipular ecuaciones algebraicas lineales

Planteamiento del problema Explore cómo MATLAB se puede emplear para resolver y analizar ecuaciones algebraicas lineales. Use el mismo sistema que en el ejemplo 11.4.

Solución Primero, introducimos la matriz $[A]$ y el vector $\{B\}$,

```
>> A = [ 1 1/2 1/3 ; 1 2/3 2/4 ; 1 3/4 3/5 ]

A =
   1.0000   0.5000   0.3333
   1.0000   0.6667   0.5000
   1.0000   0.7500   0.6000

>> B=[1+1/2+1/3;1+2/3+2/4;1+3/4+3/5]

B =
   1.8333
   2.1667
   2.3500
```

Después, se determina el número de condición para $[A]$

```
>> cond(A)

ans =
  366.3503
```

SOFTWARE

Tabla 11.1 Funciones de MATLAB para el análisis matricial y el álgebra lineal numérica.

Análisis matricial		Ecuaciones lineales	
Función	**Descripción**	**Función**	**Descripción**
cond	Número de condición de una matriz	\and/	Solución de una ecuación lineal; use "help slash"
norm	Norma vectorial o matricial	chol	Factorización de Cholesky
rcond	Estimador de condición recíproca LINPACK	lu	Factores para la eliminación de Gauss
rank	Número de renglones o columnas linealmente independientes	inv	Matriz inversa
det	Determinante	qr	Descomposición ortogonal-triangular
trace	Suma de los elementos en la diagonal	qrdelete	Suprime una columna de la factorización QR
null	Espacio nulo	qrinsert	Inserte una columna en la factorización QR
orth	Ortogonalización	nnls	Mínimos cuadrados no negativos
rref	Forma escalonada reducida por renglones	pinv	Pseudoinversa
		lscov	Mínimos cuadrados en la presencia de covarianza conocida

Este resultado se basa en la norma espectral, o $\|A\|_2$, que se analizó en el cuadro 10.2. Observe que es del mismo orden de magnitud que el número de condición = 451.2, basado en la norma renglón-suma del ejemplo 10.3. Ambos resultados implican que se podrían perder entre 2 y 3 dígitos de precisión.

Ahora se puede resolver el sistema de ecuaciones en dos formas diferentes. La forma más directa y eficiente consiste en emplear el símbolo \, o "división izquierda":

```
>> X=A\B

X =
   1.0000
   1.0000
   1.0000
```

Como en los casos anteriores, MATLAB usa la eliminación de Gauss para resolver dichos sistemas.

Como una alternativa, se puede resolver la ecuación (PT3.6) en forma directa, como

```
>> X=inv(A)*B

X =
   1.0000
   1.0000
   1.0000
```

Este procedimiento determina primero la matriz inversa y después ejecuta la multiplicación matricial. Por lo tanto, toma más tiempo que la operación de división izquierda.

11.3.3 Mathcad

Mathcad contiene muchas funciones especiales que manipulan vectores y matrices. Esto incluye operaciones ordinarias como producto punto, transposición de matrices, suma de matrices y multiplicación de matrices. Además, permite el cálculo de inverso de matriz, determinante, traza, diversos tipos de normas y números de condición con base en diferentes normas. También tiene varias funciones que descomponen matrices.

Se pueden resolver sistemas de ecuaciones lineales de dos maneras por Mathcad. Primero, es posible usar la inversión de matriz seguida por multiplicación por el lado derecho como se explicó en el capítulo 10. Además, Mathcad tiene una función especial que se llama **Isolve(A,b)** que está específicamente diseñada para resolver ecuaciones lineales. Usted puede usar otras funciones incorporadas para evaluar la condición de A y determinar si A es casi singular y, por lo tanto, posiblemente sujeta a errores de redondeo.

Como ejemplo, se usará **Isolve** para resolver un sistema de ecuaciones lineales. Como se muestra en la figura 11.8, el primer paso es introducir los coeficientes de la matriz A, usando el símbolo de definición y el menú desplegable Insert/Matrix. Esto da una caja de diálogo que le permite especificar las dimensiones de la matriz. Para el presente caso, se seleccionará una dimensión de 4 × 4, y Mathcad coloca una matriz en blanco de tamaño 4 × 4 en la pantalla. Ahora simplemente haga clic en la ubicación de la celda adecuada e introduzca los valores. Repita operaciones similares para crear el vector del lado derecho *b*. Ahora, el vector x se define como **Isolve(A,b)**, y el valor de x se muestra con el signo de igual.

También se puede resolver el mismo sistema usando la inversa de la matriz. La inversa se puede simplemente calcular con sólo elevar A al exponente –1. El resultado se muestra en el lado derecho de la figura 11.8. La solución se genera entonces como el producto de la inversa por b.

En seguida, se va a usar Mathcad para encontrar la inversa y el número de condición de la matriz de Hilbert. Como en la figura 11.9, la matriz escalada se puede introducir usando el símbolo de definición y el menú desplegable Insert/Matrix. Nuevamente, es posible calcular la inversa simplemente elevando H al exponente –1. El resultado se muestra en la figura 11.9. Se pueden entonces usar algunas otras funciones de Mathcad para determinar números de condición usando el símbolo de definición para definir las variables c1, c2, ce y ci como número de condición, basado en las normas de suma de columnas (**cond1**), espectral (**cond2**), euclidiana (**conde**) y suma de hileras (**condi**), respectivamente. Los valores resultantes se muestran en la parte inferior de la figura 11.9. Como se esperaría, la norma espectral proporciona la medición más pequeña de la magnitud.

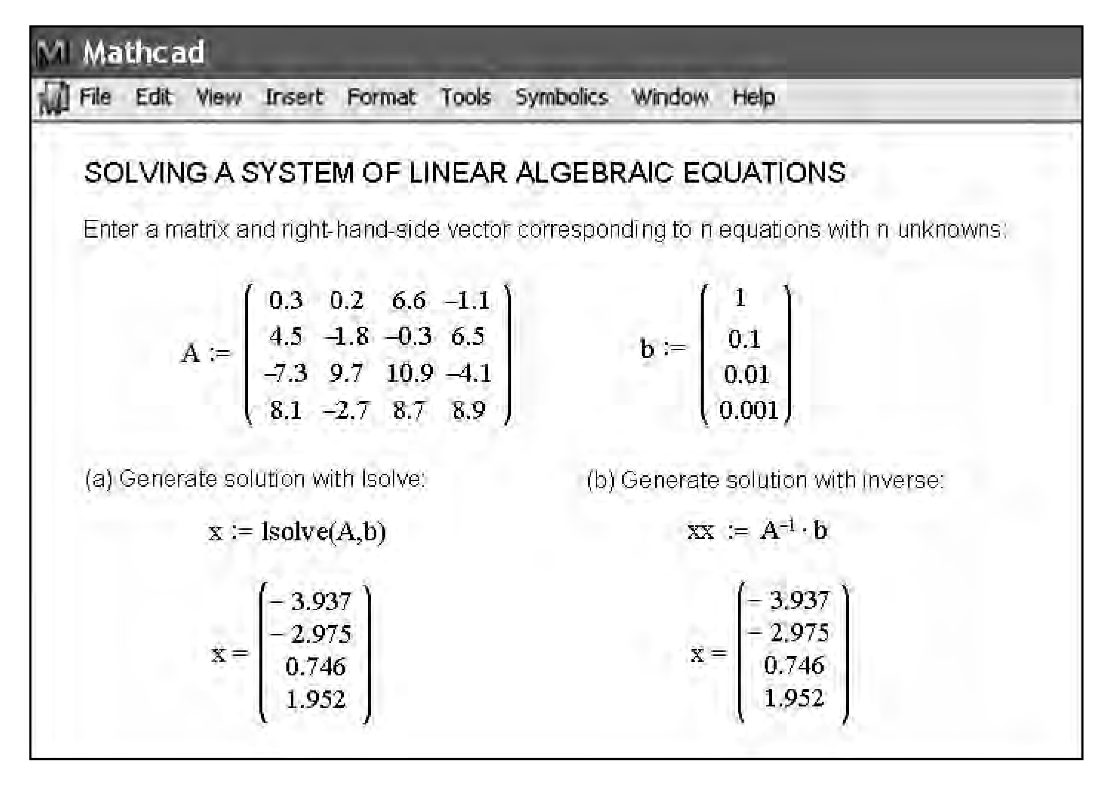


Figura 11.8
Pantalla Mathcad para resolver un sistema de ecuaciones algebraicas lineales.

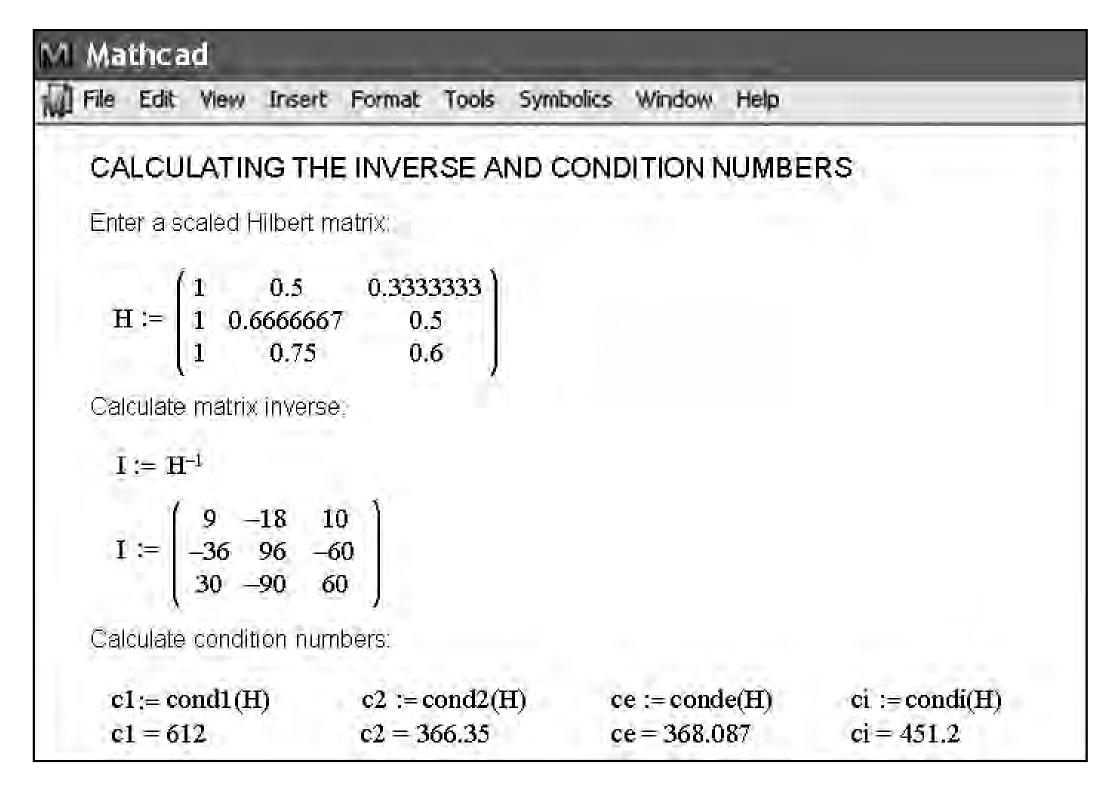


Figura 11.9
Pantalla Mathcad para determinar la inversa de la matriz y números de condición de una matriz Hilbert escalada de 3 × 3.

PROBLEMAS

11.1 Realice los mismos cálculos como en *a*) ejemplo 11.1 y *b*) ejemplo 11.3, pero para el sistema tridiagonal.

$$\begin{bmatrix} 0.8 & -0.4 & \\ -0.4 & 0.8 & -0.4 \\ & -0.4 & 0.8 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 41 \\ 25 \\ 105 \end{Bmatrix}$$

11.2 Determine la matriz inversa del ejemplo 11.1 con base en la descomposición *LU* y los vectores unitarios.

11.3 El sistema tridiagonal que sigue debe resolverse como parte de un algoritmo mayor (Crank-Nicolson) para solucionar ecuaciones diferenciales parciales:

$$\begin{bmatrix} 2.01475 & -0.020875 & & \\ -0.020875 & 2.01475 & -0.020875 & \\ & -0.020875 & 2.01475 & -0.020875 \\ & & -0.020875 & 2.01475 \end{bmatrix} \times \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{Bmatrix} = \begin{Bmatrix} 4.175 \\ 0 \\ 0 \\ 2.0875 \end{Bmatrix}$$

Utilice el algoritmo de Thomas para obtener una solución.

11.4 Confirme la validez de la descomposición de Cholesky del ejemplo 11.2 por medio de sustituir los resultados en la ecuación (11.2) con objeto de ver si el producto de $[L]$ y $[L]^T$ da como resultado $[A]$.

11.5 Haga los mismos cálculos que en el ejemplo 11.2, pero para el sistema simétrico que sigue:

$$\begin{bmatrix} 6 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} 152.6 \\ 585.6 \\ 2\,488.8 \end{Bmatrix}$$

Además de resolver para la descomposición de Cholesky, empléela para solucionar cuál es el valor de las *a*.

11.6 Ejecute a mano la descomposición de Cholesky del sistema simétrico siguiente:

$$\begin{bmatrix} 8 & 20 & 15 \\ 20 & 80 & 50 \\ 15 & 50 & 60 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 100 \\ 250 \\ 100 \end{Bmatrix}$$

11.7 Calcule la descomposición de Cholesky de

$$[A] = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 25 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

¿Su resultado tiene sentido en términos de las ecuaciones (11.3) y (11.4)?

11.8 Use el método de Gauss-Seidel para resolver el sistema tridiagonal del problema 11.1 ($\varepsilon_s = 5\%$). Utilice sobrerrelajación con $\lambda = 1.2$.

11.9 Del problema 10.8, recuerde que el sistema de ecuaciones siguiente está diseñado para determinar concentraciones (las c están en g/m^3) en una serie de reactores acoplados como función de la cantidad de masa de entrada a cada uno de ellos (los lados derechos están en g/d),

$$15c_1 - 3c_2 - c_3 = 3\,800$$
$$-3c_1 + 18c_2 - 6c_3 = 1\,200$$
$$-4c_1 - c_2 + 12c_3 = 2\,350$$

Resuelva este problema con el método de Gauss-Seidel para $\varepsilon_s = 5\%$.

11.10 Repita el problema 11.9, pero use la iteración de Jacobi.

11.11 Emplee el método de Gauss-Seidel para resolver el sistema siguiente hasta que el error relativo porcentual esté por debajo de $\varepsilon_s = 5\%$,

$$10x_1 + 2x_2 - x_3 = 27$$
$$-3x_1 - 6x_2 + 2x_3 = -61.5$$
$$x_1 + x_2 + 5x_3 = -21.5$$

11.12 Utilice el método de Gauss-Seidel *a*) sin relajación y *b*) con relajación ($\lambda = 0.95$), para resolver el sistema siguiente para una tolerancia de $\varepsilon_s = 5\%$. Si es necesario, reacomode las ecuaciones para lograr convergencia.

$$-3x_1 + x_2 + 12x_3 = 50$$
$$6x_1 - x_2 - x_3 = 3$$
$$6x_1 + 9x_2 + x_3 = 40$$

11.13 Use el método de Gauss-Seidel *a*) sin relajación y *b*) con relajación ($\lambda = 1.2$), para resolver el sistema siguiente para una tolerancia de $\varepsilon_s = 5\%$. Si es necesario, reacomode las ecuaciones para lograr convergencia.

$$2x_1 - 6x_2 - x_3 = -38$$
$$-3x_1 - x_2 + 7x_3 = -34$$
$$-8x_1 + x_2 - 2x_3 = -20$$

11.14 Vuelva a dibujar la figura 11.5 para el caso en que las pendientes de las ecuaciones son 1 y –1. ¿Cuál es el resultado de aplicar el método de Gauss-Seidel a un sistema como ése?

11.15 De los tres conjuntos siguientes de ecuaciones lineales, identifique aquel(los) que no podría resolver con el uso de un método iterativo tal como el de Gauss-Seidel. Demuestre que su solución no converge, utilizando cualquier número de iteraciones que sea necesario. Enuncie con claridad su criterio de convergencia (es decir, cómo se sabe que no está convergiendo).

Conjunto uno	Conjunto dos	Conjunto tres
$8x + 3y + z = 12$	$x + y + 5z = 7$	$-x + 3y + 5z = 7$
$-6x + 7z = 1$	$x + 4y - z = 4$	$-2x + 4y - 5z = -3$
$2x + 4y - z = 5$	$3x + y - z = 4$	$2y - z = 1$

11.16 Emplee el paquete de software de su preferencia para obtener una solución, calcular la inversa y determinar el número de condición (sin dar escala) con base en la norma de suma de renglones, para los sistemas

a)
$$\begin{bmatrix} 1 & 4 & 9 \\ 4 & 9 & 16 \\ 9 & 16 & 25 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 14 \\ 29 \\ 50 \end{Bmatrix}$$

b)
$$\begin{bmatrix} 1 & 4 & 9 & 16 \\ 4 & 9 & 16 & 25 \\ 9 & 16 & 25 & 36 \\ 16 & 25 & 36 & 49 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{Bmatrix} = \begin{Bmatrix} 30 \\ 54 \\ 86 \\ 126 \end{Bmatrix}$$

En ambos casos, las respuestas para todas las x deben ser 1.

11.17 Dado el par siguiente de ecuaciones simultáneas no lineales:

$$f(x, y) = 4 - y - 2x^2$$
$$g(x, y) = 8 - y^2 - 4x$$

a) Use la herramienta Solver de Excel para determinar los dos pares de valores de x y y que satisfacen estas ecuaciones.

b) Con el empleo de un rango de valores iniciales ($x = -6$ a 6, y $y = -6$ a 6), determine cuáles valores iniciales producen cada una de las soluciones.

11.18 Una compañía de electrónica produce transistores, resistores y chips de computadora. Cada transistor requiere cuatro unidades de cobre, una de zinc y dos de vidrio. Cada resistor requiere tres, tres y una unidades de dichos materiales, respectivamente, y cada chip de computadora requiere dos, una y tres unidades de los materiales, respectivamente. En forma de tabla, esta información queda así:

Componente	Cobre	Zinc	Vidrio
Transistores	4	1	2
Resistores	3	3	1
Chips de computadora	2	1	3

Los suministros de estos materiales varían de una semana a la otra, de modo que la compañía necesita determinar una corrida de producción diferente cada semana. Por ejemplo, cierta semana las cantidades disponibles de los materiales son 960 unidades de cobre, 510 unidades de zinc y 610 unidades de vidrio. Configure el sistema de ecuaciones que modela la corrida de producción y use software Excel, MATLAB o Mathcad para despejar el número de transistores, resistores y chips de computadora que se deben fabricar esta semana.

11.19 Utilice el software de MATLAB o Mathcad para determinar el número de condición espectral para una matriz de Hilbert de dimensión 10. ¿Cuántos dígitos de precisión se espera que se pierdan debido al mal acondicionamiento? Determine la solución para este sistema para el caso en que cada elemento del vector del lado derecho $\{b\}$ consiste en la suma de los coeficientes de su renglón. En otras palabras, resuelva para el caso en que todas las incógnitas deben ser exactamente uno. Compare los errores resultantes con aquellos esperados con base en el número de condición.

11.20 Repita el problema 11.19, pero para el caso de una matriz de Vandermonde de seis dimensiones (véase el problema 10.17) donde $x_1 = 4$, $x_2 = 2$, $x_3 = 7$, $x_4 = 10$, $x_5 = 3$ y $x_6 = 5$.

11.21 Dada una matriz cuadrada $[A]$, escriba un comando MATLAB de un solo renglón que cree una nueva matriz $[Aug]$ que consista en la matriz original $[A]$ aumentada por una matriz de identidad $[I]$.

11.22 Escriba el siguiente sistema de ecuaciones en forma matricial:

$$50 = 5x_3 - 7x_2$$
$$4x_2 + 7x_3 + 30 = 0$$
$$x_1 - 7x_3 = 40 - 3x_2 + 5x_1$$

Use software Excel, MATLAB o Mathcad para despejar las incógnitas. Además, calcule la transposición y la inversa del matriz de coeficientes.

11.23 En la sección 9.2.1 se determinó el número de operaciones que se requiere para la eliminación de Gauss sin pivoteo parcial. Efectúe una determinación similar para el algoritmo de Thomas (véase la figura 11.2). Desarrolle una gráfica de operaciones *versus* *n* (de 2 a 20) para ambas técnicas.

11.24 Desarrolle un programa amigable para el usuario en cualquier lenguaje de alto nivel o de macros, de su elección, para obtener una solución para un sistema tridiagonal con el algoritmo de Thomas (figura 11.2). Pruebe su programa por medio de repetir los resultados del ejemplo 11.1.

11.25 Desarrolle un programa amigable para el usuario en cualquier lenguaje de alto nivel o de macros, que elija, para hacer la descomposición de Cholesky con base en la figura 11.3. Pruebe su programa por medio de repetir los resultados del ejemplo 11.2.

11.26 Desarrolle un programa amigable para el usuario en cualquier lenguaje de alto nivel o de macros, que escoja, a fin de ejecutar el método de Gauss-Seidel con base en la figura 11.6. Pruébelo con la repetición de los resultados del ejemplo 11.3.

11.27 Como se describe en la sección PT3.1.2, pueden presentarse las ecuaciones algebraicas lineales en la solución de ecuaciones diferenciales. Por ejemplo, la siguiente ecuación diferencial es el resultado de un balance de masas de estado estable para un producto químico en un canal unidimensional:

$$0 = D\frac{d^2c}{dx^2} - U\frac{dc}{dx} - kc$$

donde c = concentración, t = tiempo, x = distancia, D = coeficiente de difusión, U = velocidad del fluido y k = tasa de decaimiento de primer orden. Convierta esta ecuación diferencial a un sistema equivalente de ecuaciones algebraicas simultáneas. Dados $D = 2$, $U = 1$, $k = 0.2$, $c(0) = 80$ y $c(10) = 20$, resuelva estas ecuaciones desde $x = 0$ a 10, con $\Delta x = 2$ y desarrolle una gráfica de concentración contra distancia.

11.28 Un sistema *pentadiagonal* con un ancho de banda de cinco se puede expresar generalmente como

$$\begin{bmatrix} f_1 & g_1 & h_1 & & & & \\ e_2 & f_2 & g_2 & h_2 & & & \\ d_3 & e_3 & f_3 & g_3 & h_3 & & \\ & & \cdot & \cdot & \cdot & & \\ & & & \cdot & \cdot & \cdot & \\ & & & & \cdot & \cdot & \cdot \\ & & & & d_{n-1} & e_{n-1} & f_{n-1} & g_{n-1} \\ & & & & & d_n & e_n & f_n \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ \cdot \\ \cdot \\ x_{n-1} \\ x_n \end{Bmatrix} = \begin{Bmatrix} r_1 \\ r_2 \\ r_3 \\ \cdot \\ \cdot \\ \cdot \\ r_{n-1} \\ r_n \end{Bmatrix}$$

Desarrolle un programa para resolver eficientemente tales sistemas sin estrategia de pivoteo, de manera similar al algoritmo que se usó para matrices tridiagonales en la sección 11.1.1. Pruébelo para el siguiente caso:

$$\begin{bmatrix} 8 & -2 & -1 & 0 & 0 \\ -2 & 9 & -4 & -1 & 0 \\ -1 & -3 & 7 & -1 & -2 \\ 0 & -4 & -2 & 12 & -5 \\ 0 & 0 & -7 & -3 & 15 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{Bmatrix} = \begin{Bmatrix} 5 \\ 2 \\ 0 \\ 1 \\ 5 \end{Bmatrix}$$

CAPÍTULO 12

Estudio de casos: ecuaciones algebraicas lineales

El propósito de este capítulo es usar los procedimientos numéricos analizados en los capítulos 9, 10 y 11 para resolver sistemas de ecuaciones algebraicas lineales, en algunas aplicaciones a la ingeniería. Dichas técnicas numéricas sistemáticas tienen significado práctico, ya que los ingenieros con mucha frecuencia se enfrentan a problemas que involucran sistemas de ecuaciones que son demasiado grandes para resolverse a mano. Los algoritmos numéricos en estas aplicaciones son particularmente adecuados para implementarse en computadoras personales.

En la *sección 12.1* se muestra cómo se emplea un balance de masa para modelar un sistema de reactores. En la *sección 12.2* se le da especial énfasis al uso de la matriz inversa para determinar las complicadas interacciones causa-efecto entre las fuerzas en los elementos de una armadura. La *sección 12.3* constituye un ejemplo del uso de las leyes de Kirchhoff para calcular las corrientes y voltajes en un circuito con resistores. Por último, la *sección 12.4* es una ilustración de cómo se emplean las ecuaciones lineales para determinar la configuración en estado estacionario de un sistema masa-resorte.

12.1 ANÁLISIS EN ESTADO ESTACIONARIO DE UN SISTEMA DE REACTORES (INGENIERÍA QUÍMICA/BIOINGENIERÍA)

Antecedentes Uno de los principios de organización más importantes en la ingeniería química es la *conservación de la masa* (recuerde la tabla 1.1). En términos cuantitativos, el principio se expresa como un balance de masa que toma en cuenta todas las fuentes y sumideros de un material que entra y sale de un volumen (figura 12.1). En un periodo finito, esto se expresa como

$$\text{Acumulación} = \text{entradas} - \text{salidas} \tag{12.1}$$

El balance de masa representa un ejercicio de registro para la sustancia en particular que se modela. Para el periodo en que se calcula, si las entradas son mayores que las salidas, la masa de la sustancia dentro del volumen aumenta. Si las salidas son mayores que las entradas, la masa disminuye. Si las entradas son iguales a las salidas, la acumulación es cero y la masa permanece constante. Para esta condición estable, o en estado estacionario, la ecuación (12.1) se expresa como

$$\text{Entradas} = \text{salidas} \tag{12.2}$$

Emplee la conservación de la masa para determinar las concentraciones en estado estacionario de un sistema de reactores conectados.

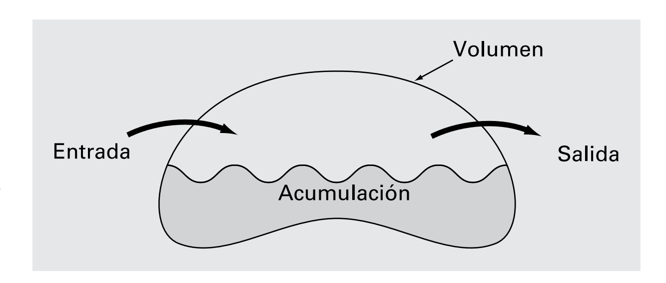


Figura 12.1
Una representación esquemática del balance de masa.

Solución Se puede usar el balance de masa para resolver problemas de ingeniería al expresar las entradas y salidas en términos de variables y parámetros medibles. Por ejemplo, si se realiza un balance de masa para una sustancia conservativa (es decir, aquella que no aumente ni disminuya debido a transformaciones químicas) en un reactor (figura 12.2), podríamos cuantificar la velocidad con la cual el flujo de la masa entra al reactor a través de dos tuberías de entrada y sale de éste a través de una tubería de salida. Esto se hace mediante el producto de la velocidad del flui-

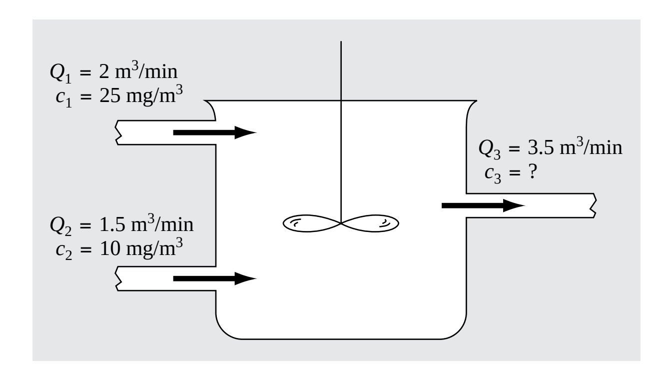


Figura 12.2
Un reactor en estado estacionario, completamente mezclado, con dos tuberías de entrada y una de salida. Los caudales *Q* están en metros cúbicos por minuto, y las concentraciones *c* están en miligramos por metro cúbico.

do o caudal Q (en metros cúbicos por minuto) por la concentración c (en miligramos por metro cúbico) en cada tubería. Por ejemplo, en la tubería 1 de la figura 12.2, $Q_1 = 2\ \text{m}^3/\text{min}$ y $c_1 = 25\ \text{mg/m}^3$; por lo tanto, la velocidad con la cual la masa fluye hacia el reactor a través de la tubería 1 es $Q_1c_1 = (2\ \text{m}^3/\text{min})(25\ \text{mg/m}^3) = 50\ \text{mg/min}$. Así, 50 mg de sustancias químicas fluyen cada minuto hacia el interior del reactor a través de esta tubería. De forma similar, para la tubería 2 la velocidad de masa que entra se calcula como $Q_2c_2 = (1.5\ \text{m}^3/\text{min})\ (10\ \text{mg/m}^3) = 15\ \text{mg/min}$.

Observe que la concentración a la salida del reactor a través de la tubería 3 no se especifica en la figura 12.2. Esto es así porque ya se tiene información suficiente para calcularla con base en la conservación de la masa. Como el reactor se halla en estado estacionario se aplica la ecuación (12.2) y las entradas deberán estar en balance con las salidas,

$$Q_1c_1 + Q_2c_2 = Q_3c_3$$

Sustituyendo los valores dados en esta ecuación se obtiene

$$50 + 15 = 3.5c_3$$

de la cual se despeja $c_3 = 18.6\ \text{mg/m}^3$. De esta forma, hemos determinado la concentración en la tercera tubería. Sin embargo, del cálculo se obtiene algo más. Como el reactor está bien mezclado (representado por el agitador en la figura 12.2), la concentración será uniforme, u homogénea, en todo el tanque. Por lo que la concentración en la tubería 3 deberá ser idéntica a la concentración en todo el reactor. En consecuencia, el balance de masa nos ha permitido calcular tanto la concentración en el reactor como en el tubo de salida. Esta información es de gran utilidad para los ingenieros químicos y petroleros, quienes tienen que diseñar reactores que tengan mezclas de una concentración específica.

Debido a que se utilizó álgebra simple para determinar la concentración de un solo reactor en la figura 12.2, podría no ser obvio lo que tiene que hacer una computadora en el cálculo de un balance de masa. En la figura 12.3 se muestra un problema donde las computadoras no solamente son útiles, sino que son de una enorme necesidad práctica. Debido a que hay cinco reactores interconectados o acoplados, se necesitan cinco ecuaciones de balance de masa para caracterizar el sistema. En el reactor 1, la velocidad de la masa que entra es

$$5(10) + Q_{31}c_3$$

y la velocidad de la masa que sale es

$$Q_{12}c_1 + Q_{15}c_1$$

Como el sistema se encuentra en estado estacionario, los flujos de entrada y de salida deben ser iguales:

$$5(10) + Q_{31}c_3 = Q_{12}c_1 + Q_{15}c_1$$

o, sustituyendo los valores de la figura 12.3,

$$6c_1 - c_3 = 50$$

Ecuaciones similares se obtienen para los otros reactores:

$$-3c_1 + 3c_2 = 0$$
$$-c_2 + 9c_3 = 160$$
$$-c_2 - 8c_3 + 11c_4 - 2c_5 = 0$$
$$-3c_1 - c_2 + 4c_5 = 0$$

Se puede utilizar un método numérico para resolver estas cinco ecuaciones con las cinco incógnitas que son las concentraciones:

$$\{C\}^T = \lfloor 11.51 \quad 11.51 \quad 19.06 \quad 17.00 \quad 11.51 \rfloor$$

Figura 12.3
Cinco reactores conectados por tuberías.

Además, la matriz inversa se calcula como

$$[A]^{-1} = \begin{bmatrix} 0.16981 & 0.00629 & 0.01887 & 0 & 0 \\ 0.16981 & 0.33962 & 0.01887 & 0 & 0 \\ 0.01887 & 0.03774 & 0.11321 & 0 & 0 \\ 0.06003 & 0.07461 & 0.08748 & 0.09091 & 0.04545 \\ 0.16981 & 0.08962 & 0.01887 & 0 & 0.25000 \end{bmatrix}$$

Cada uno de los elementos a_{ij} significa el cambio en la concentración del reactor i debido a un cambio unitario en la carga del reactor j. De esta forma, los ceros en la columna 4 indican que una carga en el reactor 4 no influirá sobre los reactores 1, 2, 3 y 5. Esto es consistente con la configuración del sistema (figura 12.3), la cual indica que el flujo de salida del reactor 4 no alimenta ningún otro reactor. En cambio, las cargas en cualquiera de los tres primeros reactores afectarán al sistema completo, como se indica por la ausencia de ceros en las primeras tres columnas. Tal información es de gran utilidad para los ingenieros que diseñan y manejan sistemas como éste.

12.2 ANÁLISIS DE UNA ARMADURA ESTÁTICAMENTE DETERMINADA (INGENIERÍA CIVIL/AMBIENTAL)

Antecedentes Un problema importante en la ingeniería estructural es encontrar las fuerzas y reacciones asociadas con una armadura estáticamente determinada. En la figura 12.4 se muestra el ejemplo de una armadura.

Las fuerzas (F) representan ya sea la tensión o la compresión sobre los componentes de la armadura. Las reacciones externas (H_2, V_2 y V_3) son fuerzas que caracterizan cómo interactúa dicha estructura con la superficie de soporte. El apoyo fijo en el nodo 2 puede transmitir fuerzas horizontales y verticales a la superficie, mientras que el apoyo móvil en el nodo 3 transmite sólo fuerzas verticales. Se observa que el efecto de la carga externa de 1 000 lb se distribuye entre los componentes de la armadura.

Solución Este tipo de estructura se puede describir como un conjunto de ecuaciones algebraicas lineales acopladas. Los diagramas de fuerza de cuerpo libre para cada nodo se muestran en la figura 12.5. La suma de las fuerzas en ambas direcciones, vertical y horizontal, deben ser cero en cada nodo, ya que el sistema está en reposo. Por lo tanto, para el nodo 1,

$$\Sigma F_H = 0 = -F_1 \cos 30° + F_3 \cos 60° + F_{1,h} \qquad (12.3)$$

$$\Sigma F_V = 0 = -F_1 \operatorname{sen} 30° - F_3 \operatorname{sen} 60° + F_{1,v} \qquad (12.4)$$

para el nodo 2,

$$\Sigma F_H = 0 = F_2 + F_1 \cos 30° + F_{2,h} + H_2 \qquad (12.5)$$

$$\Sigma F_V = 0 = F_1 \operatorname{sen} 30° + F_{2,v} + V_2 \qquad (12.6)$$

para el nodo 3,

$$\Sigma F_H = 0 = -F_2 - F_3 \cos 60° + F_{3,h} \qquad (12.7)$$

$$\Sigma F_V = 0 = F_3 \operatorname{sen} 60° + F_{3,v} + V_3 \qquad (12.8)$$

donde $F_{i,h}$ es la fuerza horizontal externa aplicada sobre el nodo i (se considera que una fuerza positiva va de izquierda a derecha) y $F_{i,v}$ es la fuerza vertical externa que se aplica sobre el nodo i (donde una fuerza positiva va hacia arriba). Así, en este problema, la fuerza de 1 000 lb hacia abajo en el nodo i corresponde a $F_{i,v}$ = –1 000 libras. En este caso, todas las otras $F_{i,v}$ y $F_{i,h}$ son cero.

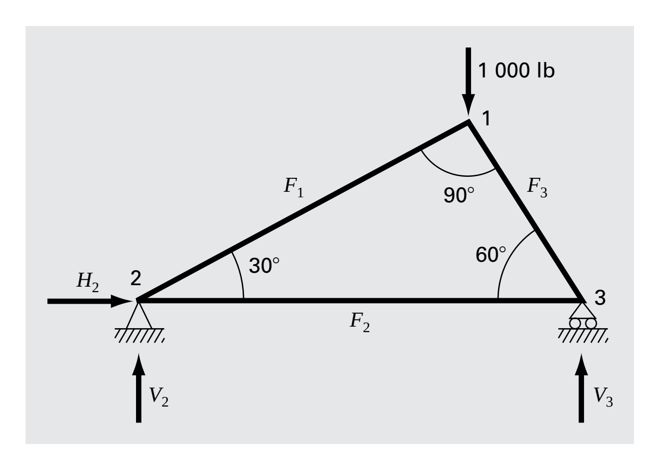


Figura 12.4
Fuerzas en una armadura estáticamente determinada.

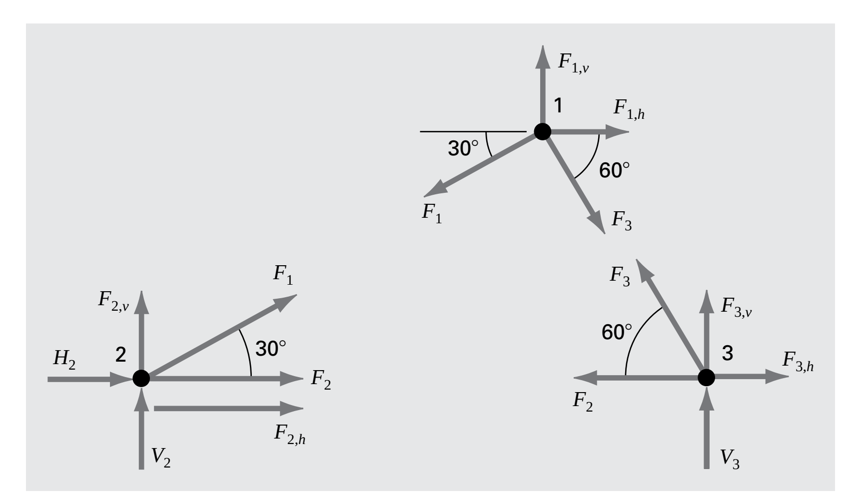


Figura 12.5
Diagramas de fuerza de cuerpo libre para los nodos de una armadura estáticamente determinada.

Observe que las direcciones de las fuerzas internas y de las reacciones son desconocidas. La aplicación correcta de las leyes de Newton requiere sólo de suposiciones consistentes respecto a la dirección. Las soluciones son negativas si las direcciones se asumen de manera incorrecta. También observe que en este problema, las fuerzas en todos los componentes se suponen en tensión y actúan tirando de los nodos adyacentes. Una solución negativa, por lo tanto, corresponde a compresión. Este problema se plantea como el siguiente sistema de seis ecuaciones con seis incógnitas:

$$\begin{bmatrix} 0.866 & 0 & -0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 0.866 & 0 & 0 & 0 \\ -0.866 & -1 & 0 & -1 & 0 & 0 \\ -0.5 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & -0.866 & 0 & 0 & -1 \end{bmatrix} \begin{Bmatrix} F_1 \\ F_2 \\ F_3 \\ H_2 \\ V_2 \\ V_3 \end{Bmatrix} = \begin{Bmatrix} 0 \\ -1\,000 \\ 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix} \tag{12.9}$$

Observe que, como se formuló en la ecuación (12.9), se requiere de pivoteo parcial para evitar la división entre cero de los elementos de la diagonal. Con el uso de una estrategia de pivote, el sistema se resuelve mediante cualquiera de las técnicas de eliminación que se analizaron en los capítulos 9 y 10. Sin embargo, como este problema es un caso de estudio ideal, para demostrar la utilidad de la matriz inversa se utiliza la descomposición *LU* para calcular

$$F_1 = -500 \qquad F_2 = 433 \qquad F_3 = -866$$
$$H_2 = 0 \qquad V_2 = 250 \qquad V_3 = 750$$

la matriz inversa es

$$[A]^{-1} = \begin{bmatrix} 0.866 & 0.5 & 0 & 0 & 0 & 0 \\ 0.25 & -0.433 & 0 & 0 & 1 & 0 \\ -0.5 & 0.866 & 0 & 0 & 0 & 0 \\ -1 & 0 & -1 & 0 & -1 & 0 \\ -0.433 & -0.25 & 0 & -1 & 0 & 0 \\ 0.433 & -0.75 & 0 & 0 & 0 & -1 \end{bmatrix}$$

Ahora, observe que el vector del lado derecho representa las fuerzas horizontales y verticales aplicadas externamente sobre cada nodo,

$$\{F\}^T = \lfloor F_{1,h} \;\; F_{1,v} \;\; F_{2,h} \;\; F_{2,v} \;\; F_{3,h} \;\; F_{3,v} \rfloor \tag{12.10}$$

Debido a que las fuerzas externas no tienen efecto sobre la descomposición *LU*, no se necesita aplicar el método una y otra vez para estudiar el efecto de diferentes fuerzas externas sobre la armadura. Todo lo que hay que hacer es ejecutar los pasos de sustitución hacia adelante y hacia atrás, para cada vector del lado derecho, y así obtener de manera eficiente soluciones alternativas. Por ejemplo, podríamos querer estudiar el efecto de fuerzas horizontales inducidas por un viento que sopla de

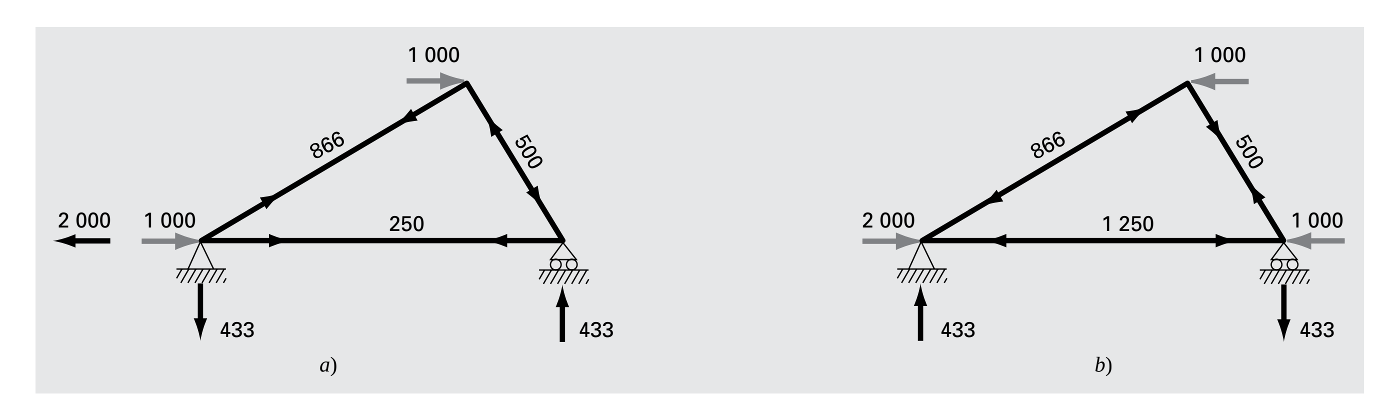


Figura 12.6
Dos casos de prueba que muestran *a*) vientos desde la izquierda y *b*) vientos desde la derecha.

izquierda a derecha. Si la fuerza del viento se puede idealizar como dos fuerzas puntuales de 1 000 libras sobre los nodos 1 y 2 (figura 12.6*a*), el vector del lado derecho es

$$\{F\}^T = \lfloor -1\,000 \quad 0 \quad 1\,000 \quad 0 \quad 0 \quad 0 \rfloor$$

que se utiliza para calcular

$$F_1 = 866 \qquad F_2 = 250 \qquad F_3 = -500$$

$$H_2 = -2\,000 \qquad V_2 = -433 \qquad V_3 = 433$$

Para un viento de la derecha (figura 12.6*b*), $F_{1,h} = -1\,000$, $F_{3,h} = -1\,000$, y todas las demás fuerzas externas son cero, con lo cual resulta

$$F_1 = -866 \qquad F_2 = -1\,250 \qquad F_3 = 500$$

$$H_2 = 2\,000 \qquad V_2 = 433 \qquad V_3 = -433$$

Los resultados indican que los vientos tienen efectos marcadamente diferentes sobre la estructura. Ambos casos se presentan en la figura 12.6.

Cada uno de los elementos de la matriz inversa tienen también utilidad directa para aclarar las interacciones estímulo-respuesta en la estructura. Cada elemento representa el cambio de una de las variables desconocidas a un cambio unitario de uno de los estímulos externos. Por ejemplo, el elemento a_{32}^{-1} indica que la tercera incógnita (F_3) cambiará a 0.866 debido a un cambio unitario del segundo estímulo externo ($F_{1,v}$). De esta forma, si la carga vertical en el primer nodo fuera aumentada en 1, F_3 se podría aumentar en 0.866. El hecho de que los elementos sean 0 indica que ciertas incógnitas no se ven afectadas por algunos de los estímulos externos. Por ejemplo, $a_{13}^{-1} = 0$ significa que F_1 no se ve afectado por cambios en $F_{2,h}$. Esta habilidad de aislar interacciones tiene diversas aplicaciones en la ingeniería; éstas comprenden la identificación de aquellos componentes que son más sensibles a estímulos externos y, como una consecuencia, más propensos a fallar. Además, esto sirve para determinar los componentes que son innecesarios (véase el problema 12.18).

El procedimiento anterior resulta particularmente útil cuando se aplica a grandes estructuras complejas. En la práctica de la ingeniería, en ocasiones es necesario resolver armaduras con cientos y aun miles de elementos estructurales. Las ecuaciones lineales proporcionan un medio poderoso para ganar cierta comprensión del comportamiento de dichas estructuras.

12.3 CORRIENTES Y VOLTAJES EN CIRCUITOS CON RESISTORES (INGENIERÍA ELÉCTRICA)

Antecedentes Un problema común en ingeniería eléctrica es la determinación de corrientes y voltajes en algunos puntos de los circuitos con resistores. Tales problemas se resuelven utilizando las reglas para corrientes y voltajes de Kirchhoff. La regla para las corrientes (o nodos) establece

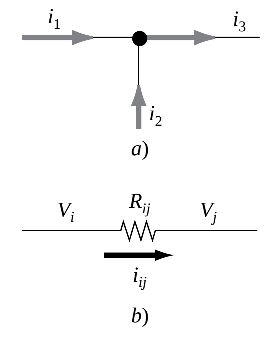


Figura 12.7
Representaciones esquemáticas de *a*) la regla de las corrientes de Kirchhoff y *b*) la ley de Ohm.

que la suma algebraica de todas las corrientes que entran a un nodo debe ser cero (véase figura 12.7*a*), o

$$\Sigma i = 0 \tag{12.11}$$

donde todas las corrientes que entran al nodo se consideran de signo positivo. La regla de las corrientes es una aplicación del principio de la conservación de la carga (recuerde la tabla 1.1).

La regla para los voltajes (o ciclos) especifica que la suma algebraica de las diferencias de potencial (es decir, cambios de voltaje) en cualquier ciclo debe ser igual a cero. Para un circuito con resistores, esto se expresa como

$$\Sigma\xi - \Sigma iR = 0 \tag{12.12}$$

donde ξ es la fem (fuerza electromotriz) de las fuentes de voltaje, y R es la resistencia de cualquier resistor en el ciclo. Observe que el segundo término se obtiene de la ley de Ohm (figura 12.7*b*), la cual establece que la caída de voltaje a través de un resistor ideal es igual al producto de la corriente por la resistencia. La regla de Kirchhoff para el voltaje es una expresión de la conservación de la energía.

Solución La aplicación de estas reglas da como resultado un sistema de ecuaciones algebraicas lineales simultáneas, ya que los ciclos que forman un circuito están conectadas. Por ejemplo, considere el circuito de la figura 12.8. Las corrientes asociadas con este circuito son desconocidas, tanto en magnitud como en dirección. Esto no presenta gran dificultad, ya que tan sólo se supone una dirección para cada corriente. Si la solución resultante a partir de las leyes de Kirchhoff es negativa, entonces la dirección supuesta fue incorrecta. Por ejemplo, la figura 12.9 muestra direcciones supuestas para las corrientes.

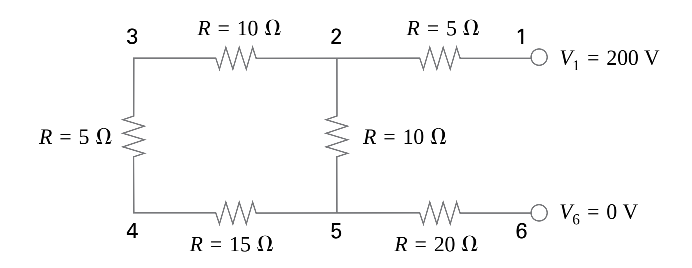


Figura 12.8
Un circuito con resistores para resolverse usando ecuaciones algebraicas lineales simultáneas.

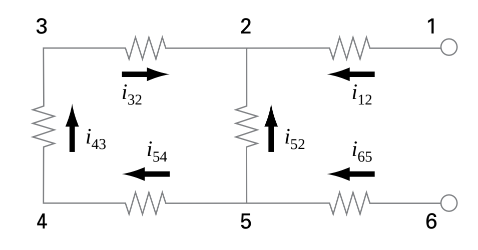


Figura 12.9
Corrientes supuestas.

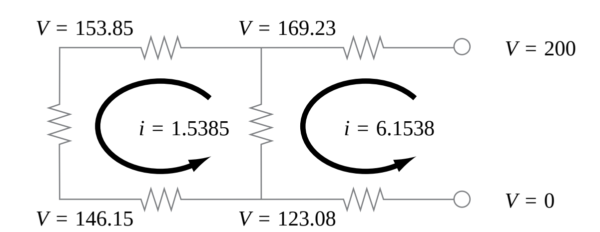


Figura 12.10
La solución obtenida para las corrientes y voltajes usando un método de eliminación.

Dadas estas suposiciones, la regla de la corriente de Kirchhoff se aplica a cada nodo para obtener

$$i_{12} + i_{52} + i_{32} = 0$$
$$i_{65} - i_{52} - i_{54} = 0$$
$$i_{43} - i_{32} = 0$$
$$i_{54} - i_{43} = 0$$

La aplicación de la regla de voltajes en cada uno de los ciclos da

$$-i_{54}R_{54} - i_{43}R_{43} - i_{32}R_{32} + i_{52}R_{52} = 0$$
$$-i_{65}R_{65} - i_{52}R_{52} + i_{12}R_{12} - 200 = 0$$

o, sustituyendo el valor de las resistencias de la figura 12.8 y pasando las constantes al lado derecho,

$$-15i_{54} - 5i_{43} - 10i_{32} + 10i_{52} = 0$$
$$-20i_{65} - 10i_{52} + 5i_{12} = 200$$

Por lo tanto, el problema consiste en la solución del siguiente conjunto de seis ecuaciones con seis corrientes como incógnitas:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 10 & -10 & 0 & -15 & -5 \\ 5 & -10 & 0 & -20 & 0 & 0 \end{bmatrix} \begin{Bmatrix} i_{12} \\ i_{52} \\ i_{32} \\ i_{65} \\ i_{54} \\ i_{43} \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 200 \end{Bmatrix}$$

Aunque no es práctico resolverlo a mano, este sistema se resuelve de manera sencilla con un método de eliminación. Si se procede de esta forma, la solución es

$$i_{12} = 6.1538 \qquad i_{52} = -4.6154 \qquad i_{32} = -1.5385$$
$$i_{65} = -6.1538 \qquad i_{54} = -1.5385 \qquad i_{43} = -1.5385$$

Así, con una interpretación adecuada de los signos del resultado, las corrientes y voltajes en el circuito se muestran en la figura 12.10. Deben ser evidentes las ventajas de usar algoritmos numéricos y computadoras para problemas de este tipo.

12.4 SISTEMAS MASA-RESORTE (INGENIERÍA MECÁNICA/AERONÁUTICA)

Antecedentes Los sistemas idealizados masa-resorte desempeñan un papel importante en la mecánica y en otros problemas de ingeniería. En la figura 12.11 se presenta un sistema de este tipo. Después de liberar las masas, éstas son jaladas hacia abajo por la fuerza de gravedad. Observe que el desplazamiento resultante en cada resorte de la figura 12.11*b* se mide a lo largo de las coordenadas locales referidas a su posición inicial en la figura 12.11*a*.

Como se mencionó en el capítulo 1, la segunda ley de Newton se emplea en conjunto con el equilibrio de fuerzas para desarrollar un modelo matemático del sistema. Para cada masa, la segunda ley se expresa como

$$m\frac{d^2x}{dt^2} = F_D - F_U \tag{12.13}$$

Para simplificar el análisis se supondrá que todos los resortes son idénticos y que se comportan de acuerdo con la ley de Hooke. En la figura 12.12*a* se muestra un diagrama de cuerpo libre para la primera masa. La fuerza hacia arriba es únicamente una expresión directa de la ley de Hooke:

$$F_U = kx_1 \tag{12.14}$$

Las componentes hacia abajo consisten en las dos fuerzas del resorte junto con la acción de la gravedad sobre la masa,

$$F_D = k(x_2 - x_1) + k(x_2 - x_1) = m_1 g \tag{12.15}$$

Observe cómo la componente de fuerza de los dos resortes es proporcional al desplazamiento de la segunda masa, x_2, corregida por el desplazamiento de la primera masa, x_1.

Las ecuaciones (12.14) y (12.15) se sustituyen en la ecuación (12.13) para dar

$$m_1\frac{d^2x_1}{dt^2} = 2k(x_2 - x_1) + m_1 g - kx_1 \tag{12.16}$$

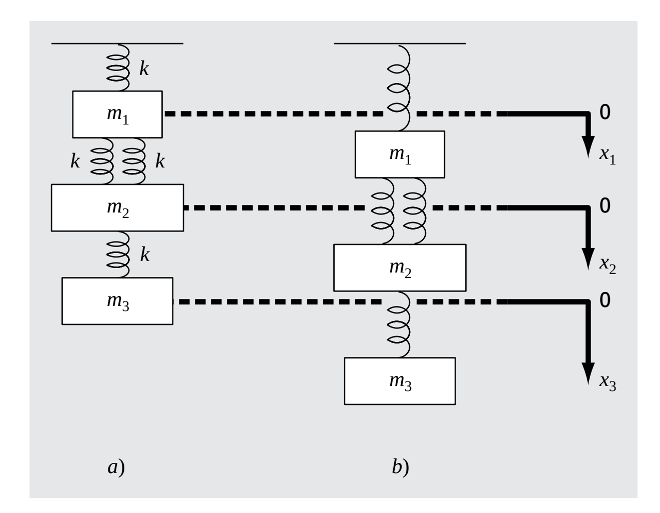


Figura 12.11
Un sistema compuesto de tres masas suspendidas verticalmente por una serie de resortes. *a*) El sistema antes de ser liberado, es decir, antes de la extensión o compresión de los resortes. *b*) El sistema después de ser liberado. Observe que las posiciones de las masas están en referencia a las coordenadas locales con orígenes en su posición antes de ser liberadas.

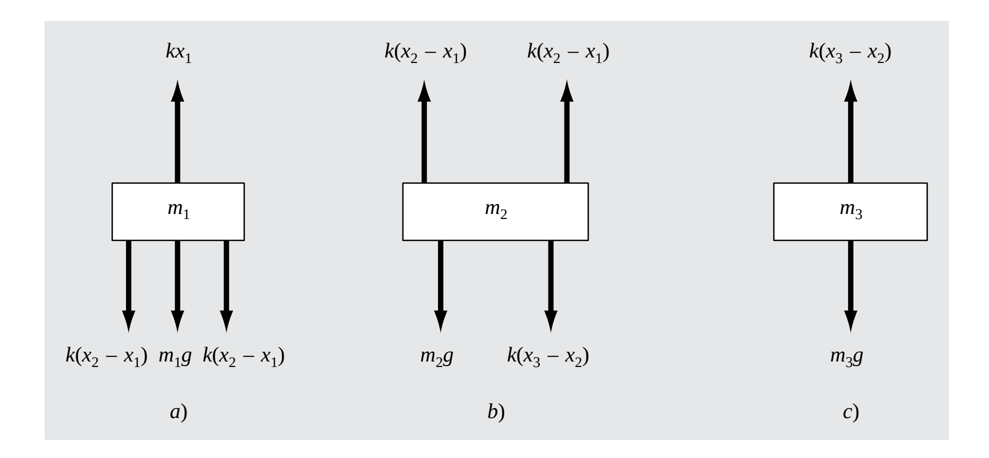


Figura 12.12
Diagramas de cuerpo libre para las tres masas de la figura 12.11.

De esta forma, se ha obtenido una ecuación diferencial ordinaria de segundo orden para describir el desplazamiento de la primera masa con respecto al tiempo. Sin embargo, advierta que la solución no se puede obtener, ya que el modelo tiene una segunda variable dependiente, x_2. En consecuencia, se deben desarrollar diagramas de cuerpo libre para la segunda y tercera masas (figuras 12.12*b* y *c*) que se emplean para obtener

$$m_2 \frac{d^2x_2}{dt^2} = k(x_3 - x_2) + m_2 g - 2k(x_2 - x_1) \tag{12.17}$$

y

$$m_3 \frac{d^2x_3}{dt^2} = m_3 g - k(x_3 - x_2) \tag{12.18}$$

Las ecuaciones (12.16), (12.17) y (12.18) forman un sistema de tres ecuaciones diferenciales con tres incógnitas. Con las condiciones iniciales apropiadas, estas ecuaciones sirven para calcular los desplazamientos de las masas como una función del tiempo (es decir, sus oscilaciones). En la parte siete estudiaremos los métodos numéricos para obtener tales soluciones. Por ahora, podemos obtener los desplazamientos que ocurren cuando el sistema eventualmente llega al reposo, es decir, al estado estacionario. Para esto se igualan a cero las derivadas en las ecuaciones (12.16), (12.17) y (12.18), obteniéndose

$$\begin{array}{rcrcrcl} 3kx_1 & - & 2kx_2 & & & = & m_1g \\ -2kx_1 & + & 3kx_2 & - & kx_3 & = & m_2g \\ & - & kx_2 & + & kx_3 & = & m_3g \end{array}$$

o, en forma matricial,

$$[K]\{X\} = \{W\}$$

donde $[K]$, conocida como *matriz de rigidez*, es

$$[K] = \begin{bmatrix} 3k & -2k & \\ -2k & 3k & -k \\ & -k & k \end{bmatrix}$$

y $\{X\}$ y $\{W\}$ son los vectores columna de las incógnitas X y de los pesos mg, respectivamente.

Solución Aquí se emplean métodos numéricos para obtener una solución. Si $m_1 = 2$ kg, $m_2 = 3$ kg, $m_3 = 2.5$ kg, y todas las $k = 10$ kg/s^2, use la descomposición *LU* con el propósito de obtener los desplazamientos y generar la inversa de $[K]$.

Sustituyendo los parámetros del modelo con g = 9.81 se obtiene

$$[K] = \begin{bmatrix} 30 & -20 & \\ -20 & 30 & -10 \\ & -10 & 10 \end{bmatrix} \qquad \{W\} = \begin{Bmatrix} 19.62 \\ 29.43 \\ 24.525 \end{Bmatrix}$$

La descomposición *LU* se utiliza con el objetivo de obtener $x_1 = 7.36$, $x_2 = 10.06$ y $x_3 = 12.51$. Estos desplazamientos se utilizaron para construir la figura 12.11*b*. La inversa de la matriz de rigidez calculada es

$$[K]^{-1} = \begin{bmatrix} 0.1 & 0.1 & 0.1 \\ 0.1 & 0.15 & 0.15 \\ 0.1 & 0.15 & 0.25 \end{bmatrix}$$

Cada elemento de la matriz k_{ji}^{-1} nos indica el desplazamiento de la masa *i* debido a una fuerza unitaria impuesta sobre la masa *j*. Así, los valores 0.1 en la columna 1 nos indican que una carga unitaria hacia abajo en la primera masa desplazará todas las masas 0.1 m hacia abajo. Los otros elementos se interpretan en forma similar. Por lo tanto, la inversa de la matriz de rigidez proporciona una síntesis de cómo los componentes del sistema responden a fuerzas que se aplican en forma externa.

PROBLEMAS

Ingeniería química/bioingeniería

12.1 Lleve a cabo el mismo cálculo que en la sección 12.1, pero cambie c_{01} a 20 y c_{03} a 6. También cambie los flujos siguientes: $Q_{01} = 6$, $Q_{12} = 4$, $Q_{24} = 2$ y $Q_{44} = 12$.

12.2 Si la entrada al reactor 3 de la sección 12.1 disminuye 25%, utilice la matriz inversa para calcular el cambio porcentual en la concentración de los reactores 2 y 4.

12.3 Debido a que el sistema que se muestra en la figura 12.3 está en estado estacionario (estable), ¿qué se puede afirmar respecto de los cuatro flujos: Q_{01}, Q_{03}, Q_{44} y Q_{55}?

12.4 Vuelva a calcular las concentraciones para los cinco reactores que se muestran en la figura 12.3, si los flujos cambian como sigue:

$Q_{01} = 5$	$Q_{31} = 3$	$Q_{25} = 2$	$Q_{23} = 2$
$Q_{15} = 4$	$Q_{55} = 3$	$Q_{54} = 3$	$Q_{34} = 7$
$Q_{12} = 4$	$Q_{03} = 8$	$Q_{24} = 0$	$Q_{44} = 10$

12.5 Resuelva el mismo sistema que se especifica en el problema 12.4, pero haga $Q_{12} = Q_{54} = 0$ y $Q_{15} = Q_{34} = 3$. Suponga que las entradas (Q_{01}, Q_{03}) y las salidas (Q_{44}, Q_{55}) son las mismas. Use la conservación del flujo para volver a calcular los valores de los demás flujos.

12.6 En la figura P12.6 se muestran tres reactores conectados por tubos. Como se indica, la tasa de transferencia de productos químicos a través de cada tubo es igual a la tasa de flujo (*Q*, en unidades de metros cúbicos por segundo) multiplicada por la concentración del reactor desde el que se origina el flujo (*c*, en unidades de miligramos por metro cúbico). Si el sistema se encuentra en estado estacionario (estable), la transferencia de entrada a cada reactor balanceará la de salida. Desarrolle las ecuaciones del balance de masa para los reactores y resuelva las tres ecuaciones algebraicas lineales simultáneas para sus concentraciones.

12.7 Con el empleo del mismo enfoque que en la sección 12.1, determine la concentración de cloruro en cada uno de los Grandes Lagos con el uso de la información que se muestra en la figura P12.7.

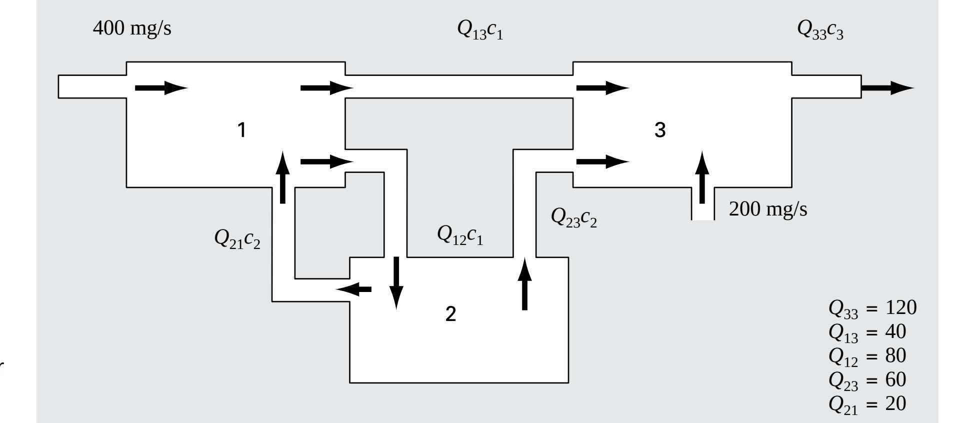


Figura P12.6
Tres reactores unidos por tubos. La tasa de transferencia de masa a través de cada tubo es igual al producto de flujo *Q* y la concentración *c* del reactor desde el que se origina el flujo.

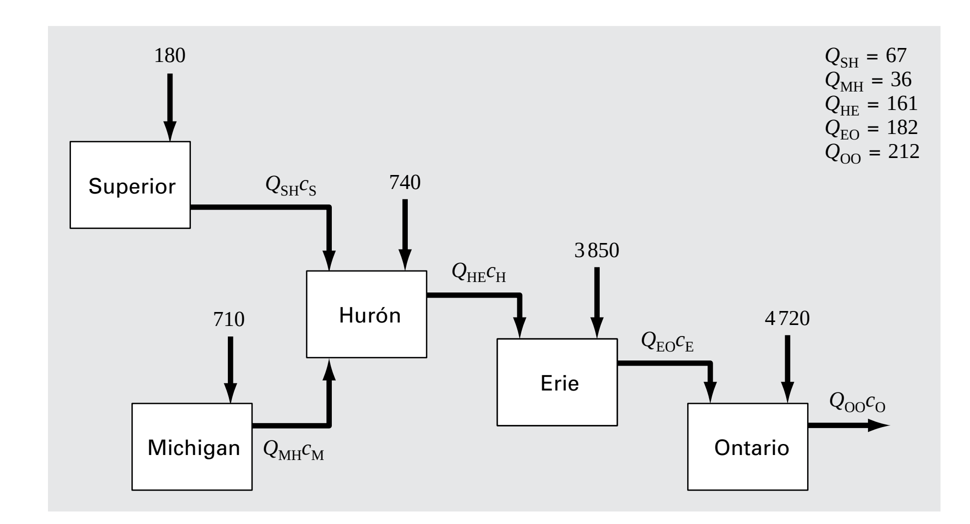


Figura P12.7
Balance del cloro en los Grandes Lagos. Las flechas numeradas denotan entradas directas.

12.8 La parte baja del río Colorado consiste en una serie de cuatro almacenamientos como se ilustra en la figura P12.8. Pueden escribirse los balances de masa para cada uno de ellos, lo que da por resultado el conjunto siguiente de ecuaciones algebraicas lineales simultáneas:

$$\begin{bmatrix} 13.422 & 0 & 0 & 0 \\ -13.422 & 12.252 & 0 & 0 \\ 0 & -12.252 & 12.377 & 0 \\ 0 & 0 & -12.377 & 11.797 \end{bmatrix} \begin{Bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{Bmatrix} = \begin{Bmatrix} 750.5 \\ 300 \\ 102 \\ 30 \end{Bmatrix}$$

donde el vector del lado derecho consiste en las cargas de cloruro hacia cada uno de los cuatro lagos y c_1, c_2, c_3 y c_4 = las concentraciones de cloruro resultantes en los lagos Powell, Mead, Mohave y Havasu, respectivamente.

a) Use la matriz inversa para resolver cuáles son las concentraciones en cada uno de los cuatro lagos.
b) ¿En cuánto debe reducirse la carga del lago Powell para que la concentración de cloruro en el lago Havasu sea de 75?
c) Con el uso de la norma columna-suma, calcule el número de condición y diga cuántos dígitos sospechosos se generarían al resolver este sistema.

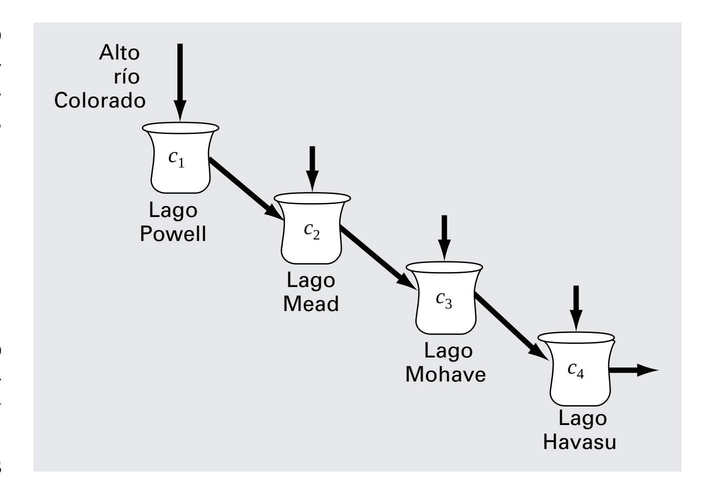


Figura P12.8
Caída del río Colorado.

12.9 En la figura P12.9 se ilustra un proceso de extracción en etapas. En tales sistemas, una corriente que contiene una fracción de peso Y_{ent} de un producto químico ingresa por la izquierda con una tasa de flujo de masa de F_1. En forma simultánea, un solvente que lleva una fracción de peso X_{ent} del mismo producto químico entra por la derecha con una tasa de flujo de F_2. Así, para la etapa i, el balance de masa se representa como

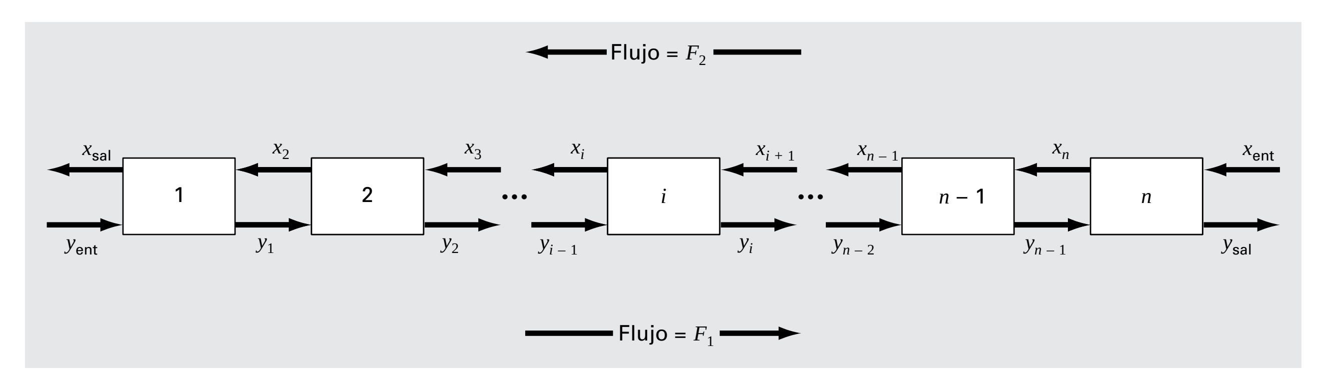


Figura P12.9
Una etapa del proceso de extracción.

$$F_1Y_{i-1} + F_2X_{i+1} = F_1Y_i + F_2X_i \qquad \text{(P12.9.1)}$$

En cada etapa, se supone que se establece el equilibrio entre Y_i y X_i, como en

$$K = \frac{X_i}{Y_i} \qquad \text{(P12.9.2)}$$

donde K se denomina coeficiente de distribución. La ecuación (P12.9.2) puede resolverse para X_i y se sustituye en la ecuación (P12.9.1) para producir

$$Y_{i-1} - \left(1 + \frac{F_2}{F_1}K\right)Y_i + \left(\frac{F_2}{F_1}K\right)Y_{i+1} = 0 \qquad \text{(P12.9.3)}$$

Si F_1 = 400 kg/h, Y_{ent} = 0.1, F_2 = 800 kg/h, X_{ent} = 0 y K = 5, determine los valores de Y_{sal} y X_{sal}, si se emplea un reactor de cinco etapas. Obsérvese que debe modificarse la ecuación (P12.9.3) para tomar en cuenta las fracciones de peso del flujo de entrada cuando se aplique a la primera y última etapas.

12.10 Una reacción de primer orden, irreversible, tiene lugar en cuatro reactores bien mezclados (véase la figura P12.10),

$$A \xrightarrow{k} B$$

Así, la tasa a la cual A se transforma en B se representa por

$$R_{ab} = kVc$$

Los reactores tienen volúmenes diferentes, y debido a que se operan a temperaturas diferentes, cada uno tiene distinta tasa de reacción:

Reactor	*V*, L	*k*, h⁻¹
1	25	0.05
2	75	0.1
3	100	0.5
4	25	0.1

Determine la concentración de A y B en cada uno de los reactores en estado estable.

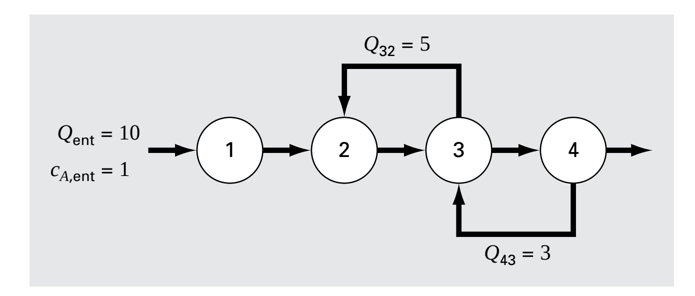


Figura P12.10

12.11 Una bomba peristáltica envía un flujo unitario (Q_1) de un fluido muy viscoso. En la figura P12.11 se ilustra la red. Cada sección de tubo tiene la misma longitud y diámetro. El balance de masa y energía mecánica se simplifica para obtener los flujos en cada tubo. Resuelva el sistema de ecuaciones siguiente a fin de obtener el flujo en cada tubo.

$$Q_3 + 2Q_4 - 2Q_2 = 0$$
$$Q_5 + 2Q_6 - 2Q_4 = 0$$
$$3Q_7 - 2Q_6 = 0$$

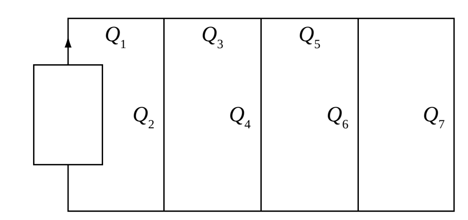


Figura P12.11

$$Q_1 = Q_2 + Q_3$$
$$Q_3 = Q_4 + Q_5$$
$$Q_5 = Q_6 + Q_7$$

12.12 La figura P12.12 ilustra un proceso de intercambio químico que consiste en una serie de reactores en los que un gas que fluye de izquierda a derecha pasa por un líquido que fluye de derecha a izquierda. La transferencia de un producto químico del gas al líquido ocurre a una tasa proporcional a la diferencia entre las concentraciones del gas y el líquido en cada reactor. En estado estacionario (estable), el balance de masa para el primer reactor se puede escribir para el gas, así

$$Q_G c_{G0} - Q_G c_{G1} + D(c_{L1} - c_{G1}) = 0$$

y para el líquido,

$$Q_L c_{L2} - Q_L c_{L1} + D(c_{G1} - c_{L1}) = 0$$

donde Q_G y Q_L son las tasas de flujo del gas y el líquido, respectivamente, y D = tasa de intercambio gas-líquido. Es posible escribir otros balances similares para los demás reactores. Resuelva para las concentraciones con los siguientes valores dados: Q_G = 2, Q_L = 1, D = 0.8, c_{G0} = 100, c_{L6} = 20.

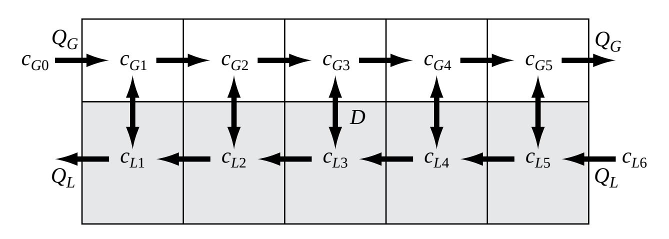


Figura P12.12

Ingeniería civil/ambiental

12.13 Un ingeniero civil que trabaja en la construcción requiere 4 800, 5 800 y 5 690 m³ de arena, grava fina y grava gruesa, respectivamente, para cierto proyecto constructivo. Hay tres canteras de las que puede obtenerse dichos materiales. La composición de dichas canteras es la que sigue

	Arena %	Grava fina %	Grava gruesa %
Cantera 1	52	30	18
Cantera 2	20	50	30
Cantera 3	25	20	55

¿Cuántos metros cúbicos deben extraerse de cada cantera a fin de satisfacer las necesidades del ingeniero?

12.14 Ejecute el mismo cálculo que en la sección 12.2, pero para la armadura que se ilustra en la figura P12.14.

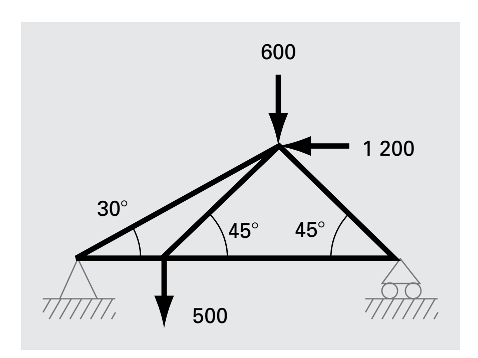


Figura P12.14

12.15 Realice el mismo cálculo que en la sección 12.2, pero para la armadura que se muestra en la figura P12.15.

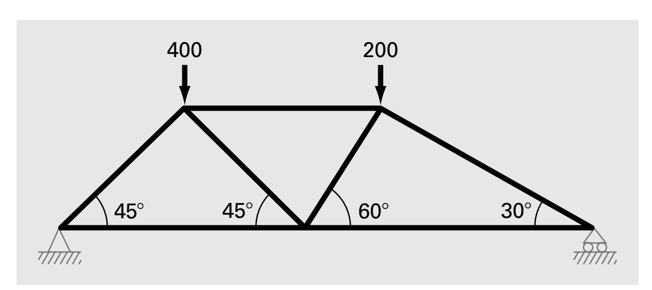


Figura P12.15

12.16 Calcule las fuerzas y reacciones para la viga de la figura 12.4, si en el nodo 1 se aplica una fuerza hacia abajo de 2 500 kg y otra horizontal hacia la derecha de 2 000 kg.

12.17 En el ejemplo de la figura 12.4, donde en el nodo 1 se aplica una fuerza debajo de 1 000 libras, se calcularon las reacciones externas V_2 y V_3. Pero si se hubieran dado las longitudes de los miembros de las armaduras habría podido calcularse V_2 y V_3 haciendo uso del hecho de que $V_2 + V_3$ debe ser igual a 1 000, y con la suma de momentos alrededor del nodo 2. Sin embargo, debido a que se conocen V_2 y V_3, es posible trabajar a la inversa para resolver cuáles son las longitudes de los miembros de la armadura. Obsérvese que debido a que hay tres longitudes desconocidas y sólo dos ecuaciones, se puede resolver sólo para la relación entre las longitudes. Resuelva para esta relación.

12.18 Con el mismo método que se usó para analizar la figura 12.4, determine las fuerzas y reacciones para la armadura que se ilustra en la figura P12.18.

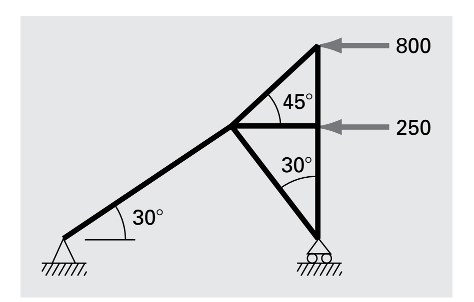


Figura P12.18

12.19 Resuelva para las fuerzas y reacciones para la armadura que se ilustra en la figura P12.19. Determine la matriz inversa para el sistema. ¿Parece razonable la fuerza del miembro vertical en el miembro de en medio? ¿Por qué?

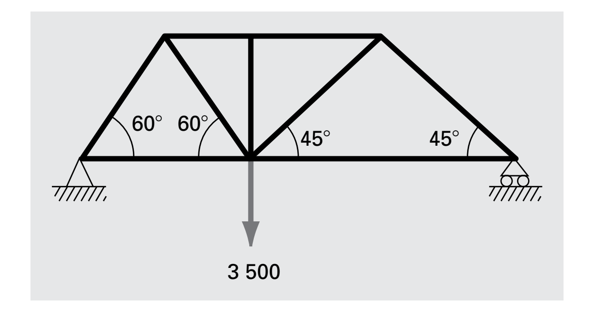


Figura P12.19

12.20 Como su nombre lo dice, la contaminación del aire interior se refiere a la contaminación del aire en espacios cerrados, como casas, oficinas, áreas de trabajo, etc. Suponga que usted está diseñando el sistema de ventilación para un restaurante como se ilustra en la figura P12.20. El área de servicio del restaurante consiste en dos habitaciones cuadradas y otra alargada. La habitación 1 y la 3 tienen fuentes de monóxido de carbono que proviene de los fumadores y de una parrilla defectuosa, respectivamente. Es posible plantear los balances de masa en estado estacionario para cada habitación. Por ejemplo, para la sección de fumadores (habitación 1), el balance es el siguiente

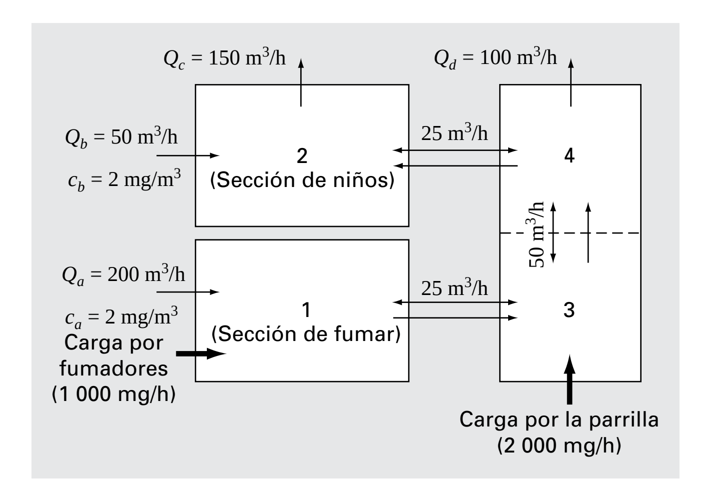


Figura P12.20
Vista de arriba de las áreas en un restaurante. Las flechas en un solo sentido representan flujos volumétricos de aire, mientras que las de dos sentidos indican mezclas difusivas. Las cargas debidas a los fumadores y a la parrilla agregan masa de monóxido de carbono al sistema pero con un flujo de aire despreciable.

$$0 = W_{\text{fumador}} + Q_a c_a - Q_a c_1 + E_{13}(c_3 - c_1)$$
$$(\text{carga}) + (\text{entrada}) - (\text{salida}) + (\text{mezcla})$$

o al sustituir los parámetros

$$225c_1 - 25c_3 = 2\,400$$

Para las demás habitaciones se pueden escribir balances similares.

a) Resuelva para la concentración de monóxido de carbono en estado estacionario en cada habitación.

b) Determine qué porcentaje del monóxido de carbono en la sección de niños se debe a *i*) los fumadores, *ii*) la parrilla y *iii*) el aire que entra por ventilación.

c) Si las cargas de los fumadores y la parrilla se incrementan a 2 000 y 5 000 mg/h, respectivamente, utilice la matriz inversa para determinar el aumento en la concentración en la sección de niños.

d) ¿Cómo cambia la concentración en el área de niños si se construye una pantalla de modo que la mezcla entre las áreas 2 y 4 disminuya a 5 m^3/h?

12.21 Se aplica una fuerza hacia arriba de 20 kN en la cúspide de un trípode como se ilustra en la figura P12.21. Determine las fuerzas en las patas del trípode.

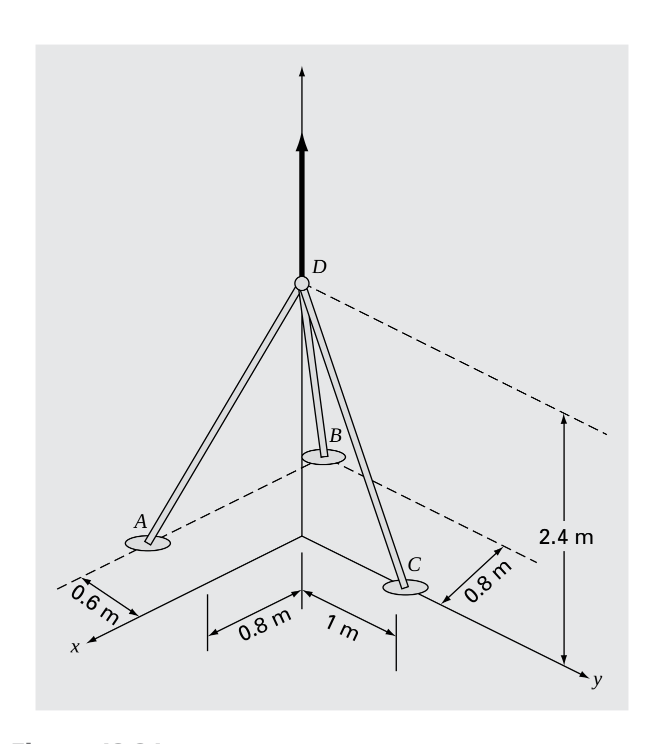


Figura 12.21

12.22 Se carga una armadura según se ilustra en la figura P12.22. Con el uso del conjunto siguiente de ecuaciones, resuelva para las 10 incógnitas: *AB*, *BC*, *AD*, *BD*, *CD*, *DE*, *CE*, A_x, A_y y E_y.

$$A_x + AD = 0 \qquad -24 - CD - (4/5)CE = 0$$
$$A_y + AB = 0 \qquad -AD + DE - (3/5)BD = 0$$
$$74 + BC + (3/5)BD = 0 \qquad CD + (4/5)BD = 0$$

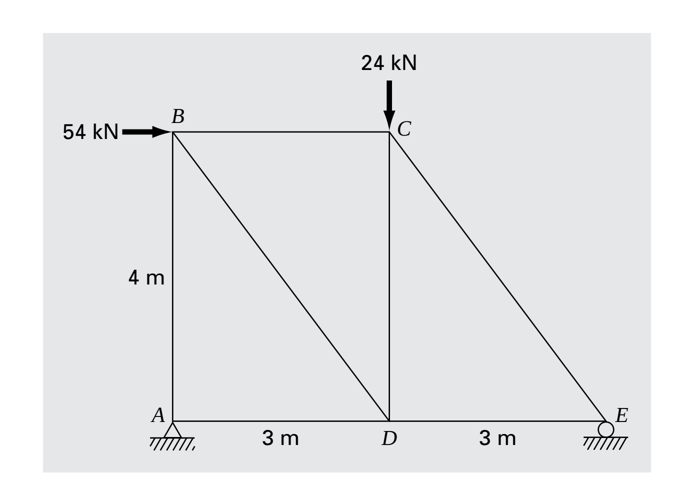


Figura P12.22

$$-AB - (4/5)BD = 0 \qquad -DE - (3/5)CE = 0$$
$$-BC + (3/5)CE = 0 \qquad E_y + (4/5)CE = 0$$

Ingeniería eléctrica

12.23 Efectúe el mismo cálculo que en la sección 12.3, pero para el circuito que se ilustra en la figura P12.23.

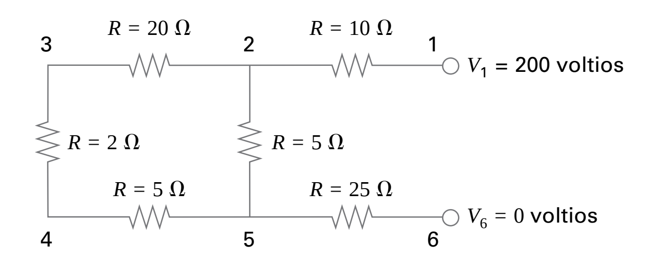


Figura P12.23

12.24 Realice el mismo cálculo que en la sección 12.3, pero para el circuito que se muestra en la figura P12.24.

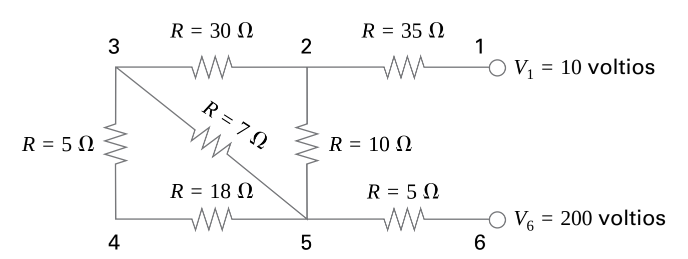


Figura P12.24

12.25 Resuelva el circuito que aparece en la figura P12.25, para las corrientes en cada conductor. Utilice la eliminación de Gauss con pivoteo.

12.26 Un ingeniero eléctrico supervisa la producción de tres tipos de componentes eléctricos. Para ello se requieren tres clases de material: metal, plástico y caucho. A continuación se presentan las cantidades necesarias para producir cada componente.

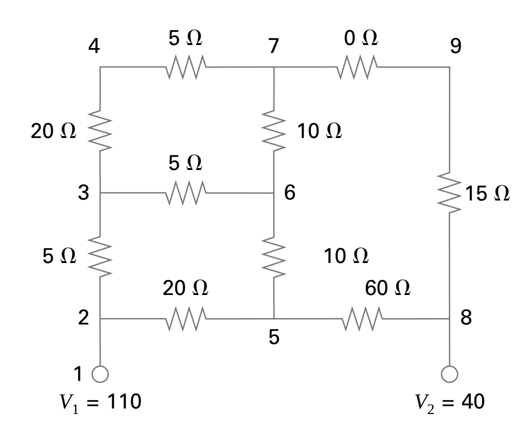


Figura P12.25

Componente	Metal, g/componente	Plástico, g/componente	Caucho g/componente
1	15	0.30	1.0
2	17	0.40	1.2
3	19	0.55	1.5

Si cada día se dispone de un total de 3.88, 0.095 y 0.282 kg de metal, plástico y caucho, respectivamente, ¿cuántos componentes pueden producirse por día?

12.27 Determine las corrientes para el circuito de la figura P12.27.

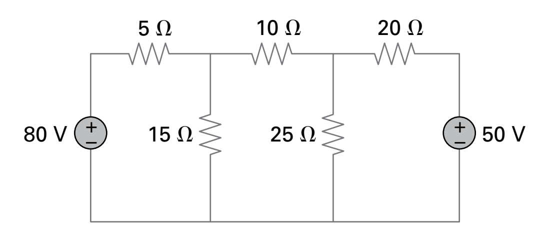


Figura P12.27

12.28 Calcule las corrientes en el circuito que aparece en la figura P12.28.

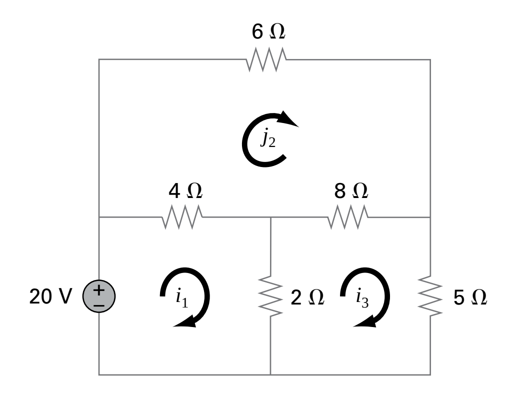


Figura P12.28

12.29 El sistema de ecuaciones que sigue se generó por medio de aplicar la ley de malla de corrientes al circuito de la figura P12.29:

$$55I_1 - 25I_4 = -200$$
$$-37I_3 - 4I_4 = -250$$
$$-25I_1 - 4I_3 + 29I_4 = 100$$

Encuentre I_1, I_3 e I_4.

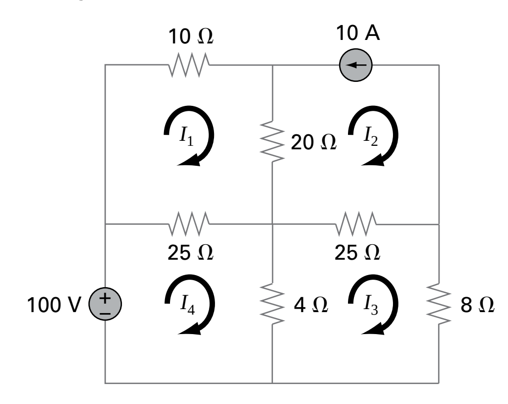


Figura P12.29

12.30 El sistema de ecuaciones siguiente se generó con la aplicación de la ley de malla de corrientes al circuito de la figura P12.30:

$$60I_1 - 40I_2 = 200$$
$$-40I_1 + 150I_2 - 100I_3 = 0$$
$$-100I_2 + 130I_3 = 230$$

Encuentre I_1, I_2 e I_3.

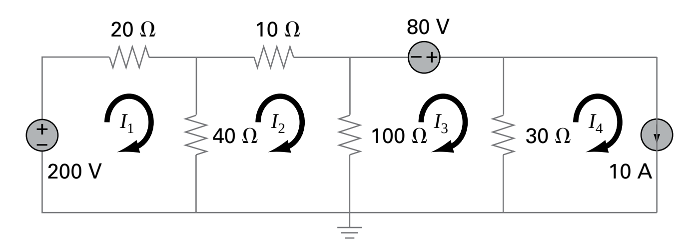


Figura P12.30

Ingeniería mecánica/aeroespacial

12.31 Lleve a cabo el mismo cálculo que en la sección 12.4, pero agregue un tercer resorte entre las masas 1 y 2, y triplique el valor de k para todos los resortes.

12.32 Realice el mismo cálculo que en la sección 12.4, pero cambie las masas de 2, 3 y 2.5 kg por otras de 10, 3.5 y 2 kg, respectivamente.

12.33 Los sistemas idealizados de masa-resorte tienen aplicaciones numerosas en la ingeniería. La figura P12.33 muestra un arreglo de cuatro resortes en serie comprimidos por una fuerza de 2000 kg. En el equilibrio, es posible desarrollar ecuaciones de balance de fuerza si se definen las relaciones entre los resortes.

$$k_2(x_2 - x_1) = k_1x_1$$
$$k_3(x_3 - x_2) = k_2(x_2 - x_1)$$
$$k_4(x_4 - x_3) = k_3(x_3 - x_2)$$
$$F = k_4(x_4 - x_3)$$

donde las k son constantes de los resortes. Si de k_1 a k_4 son 150, 50, 75 y 225 N/m, respectivamente, calcule el valor de las x.

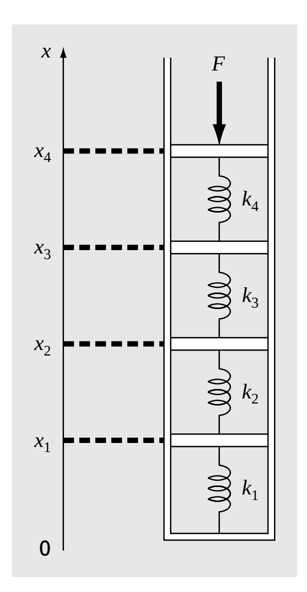


Figura P12.33

12.34 Se conectan tres bloques por medio de cuerdas carentes de peso y se dejan en reposo en un plano inclinado (véase la figura P12.34*a*). Con el empleo de un procedimiento similar al que se usó en el análisis del paracaidista en descenso del ejemplo 9.11 se llega al conjunto siguiente de ecuaciones simultáneas (en la figura P12.34*b* se muestran los diagramas de cuerpo libre):

$$
\begin{aligned}
100a + T \quad &= 519.72\\
50a - T + R &= 216.55\\
25a \quad - R &= 108.28
\end{aligned}
$$

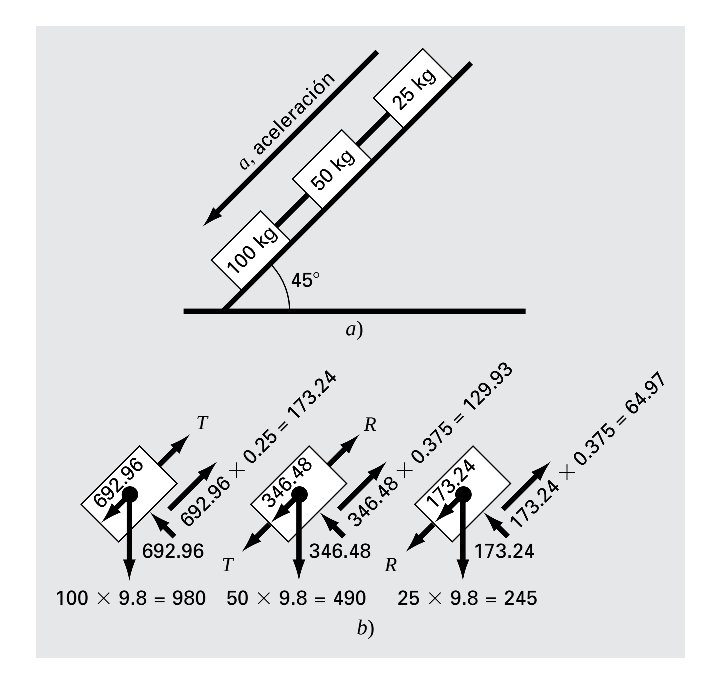


Figura P12.34

Resuelva para la aceleración a y las tensiones T y R en las dos cuerdas.

12.35 Efectúe un cálculo similar al que se utilizó en el problema P12.34, pero para el sistema que se ilustra en la figura P12.35.

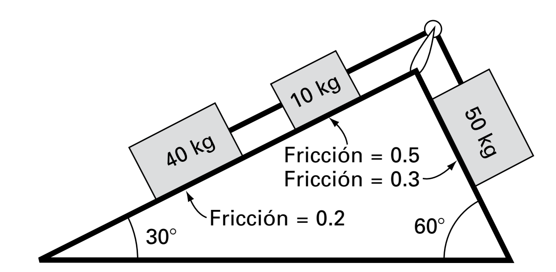


Figura P12.35

12.36 Realice el mismo cálculo que en el problema 12.34, pero para el sistema que se muestra en la figura P12.36 (los ángulos son de 45°).

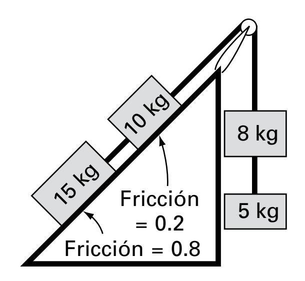


Figura P12.36

12.37 Considere el sistema de tres masas y cuatro resortes que aparece en la figura P12.37. Al determinar las ecuaciones de movimiento a partir de $\sum F_x = ma$, para cada masa con el empleo de su diagrama de cuerpo libre, se llega a las ecuaciones diferenciales siguientes:

$$\ddot{x}_1 + \left(\frac{k_1 + k_2}{m_1}\right)x_1 - \left(\frac{k_2}{m_1}\right)x_2 = 0$$

$$\ddot{x}_2 - \left(\frac{k_2}{m_2}\right)x_1 + \left(\frac{k_2 + k_3}{m_2}\right)x_2 - \left(\frac{k_3}{m_2}\right)x_3 = 0$$

$$\ddot{x}_3 - \left(\frac{k_3}{m_3}\right)x_2 + \left(\frac{k_3 + k_4}{m_3}\right)x_3 = 0$$

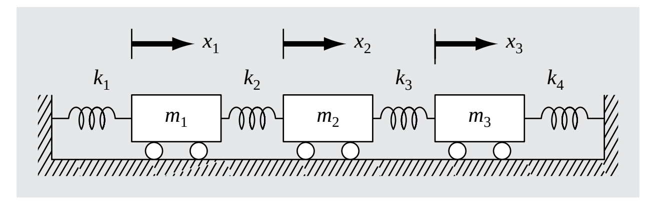


Figura P12.37

donde $k_1 = k_4 = 10$ N/m, $k_2 = k_3 = 30$ N/m y $m_1 = m_2 = m_3 = 2$ kg. Escriba las tres ecuaciones en forma matricial:

0 = [vector de aceleración] + [matriz *k*/*m*] [vector de desplazamiento *x*]

En un momento específico en el que $x_1 = 0.05$ m, $x_2 = 0.04$ m y $x_3 = 0.03$ m, se forma una matriz tridiagonal. Resuelva cuál es la aceleración de cada masa.

12.38 Las ecuaciones algebraicas lineales pueden surgir al resolver ecuaciones diferenciales. Por ejemplo, la ecuación diferencial siguiente proviene de un balance de calor para una barra larga y delgada (véase la figura P12.38):

$$\frac{d^2T}{dx^2} + h'(T_a - T) = 0 \qquad \text{(P12.38.1)}$$

donde T = temperatura (°C), x = distancia a lo largo de la barra (*m*), h' = coeficiente de transferencia de calor entre la barra y el aire del ambiente (m^{-2}) y T_a = temperatura del aire circundante (°C). Esta ecuación se transforma en un conjunto de ecuaciones algebraicas lineales por medio del uso de una aproximación en diferencias finitas divididas para la segunda derivada (recuerde la sección 4.1.3),

$$\frac{d^2T}{dx^2} = \frac{T_{i+1} - 2T_i + T_{i-1}}{\Delta x^2}$$

donde T_i denota la temperatura en el nodo *i*. Esta aproximación se sustituye en la ecuación (P12.38.1) y se obtiene

$$-T_{i-1} + (2 + h'\Delta x^2)T_i - T_{i+1} = h'\Delta x^2 T_a$$

Se puede plantear esta ecuación para cada uno de los nodos interiores de la barra, lo que resulta en un sistema tridiagonal de ecuaciones. Los nodos primero y último en los extremos de la barra están fijos por las condiciones de frontera.

a) Desarrolle la solución analítica para la ecuación (P12.38.1) para una barra de 10 m con $T_a = 20$, $T(x = 0) = 40$, $T(x = 10) = 200$ y $h' = 0.02$.

b) Desarrolle una solución numérica para los mismos valores de los parámetros que se emplearon en el inciso *a*), con el uso de una solución en diferencias finitas con cuatro nodos interiores según se muestra en la figura P12.38 ($\Delta x = 2$ m).

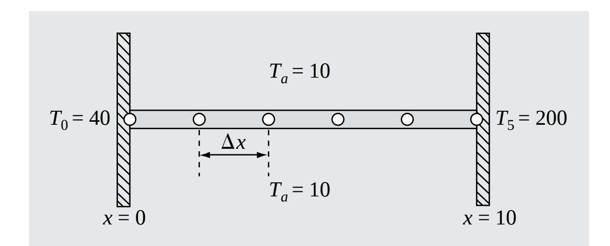


Figura P12.38
Una barra uniforme sin aislamiento colocada entre dos paredes de temperatura constante pero diferente. La representación en diferencias finitas emplea cuatro nodos interiores.

12.39 La distribución de temperatura de estado estable en una placa caliente está modelada por la *ecuación de Laplace*:

$$0 = \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2}$$

Si se representa la placa por una serie de nodos (véase la figura P12.39), las diferencias finitas divididas se pueden sustituir por las segundas derivadas, lo que da como resultado un sistema de ecuaciones algebraicas lineales. Utilice el método de Gauss-Seidel para resolver cuáles son las temperaturas de los nodos que se aprecian en la figura P12.39.

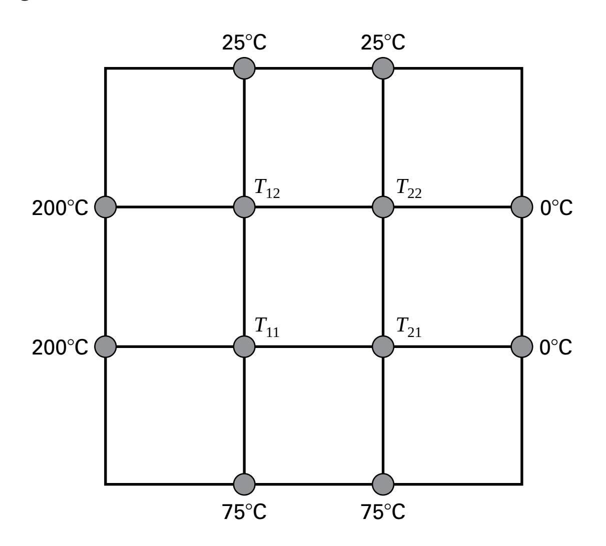


Figura P12.39

12.40 Una barra sobre una bola y una junta tipo *socket* está sujeta a los cables *A* y *B* como se observa en la figura P12.40.

a) Si se ejerce una fuerza de 50 N sobre la barra sin masa en *G*, ¿cuál es la fuerza de la tensión en los cables *A* y *B*?

b) Resuelva cuáles son las fuerzas de reacción en la base de la barra. Denomine al punto de la base como *P*.

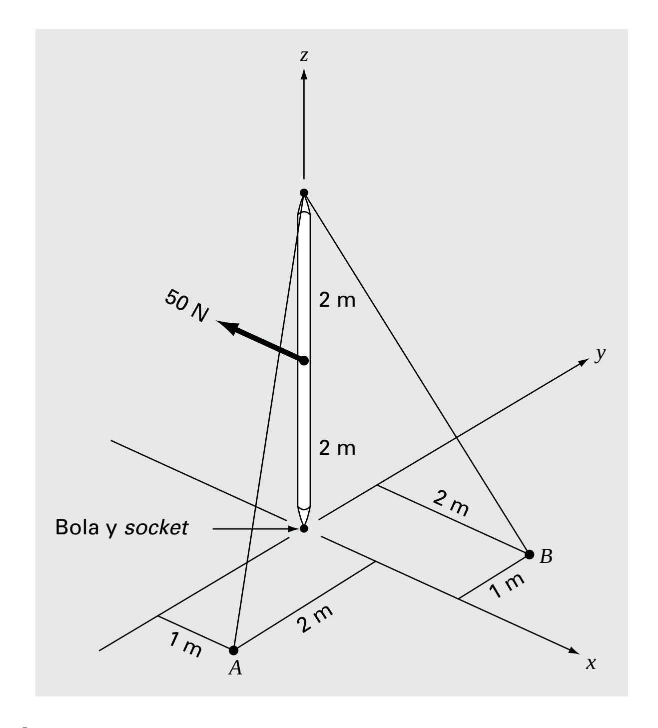


Figura P12.40

EPÍLOGO: PARTE TRES

PT3.4 ALTERNATIVAS

La tabla PT3.2 ofrece un resumen de las ventajas y desventajas en la solución de ecuaciones algebraicas lineales simultáneas. Dos métodos (el gráfico y la regla de Cramer) están limitados a pocas ecuaciones (≤ 3), de modo que tienen escasa utilidad para resolver problemas prácticos. Sin embargo, dichas técnicas son herramientas didácticas útiles para entender el comportamiento de los sistemas lineales en general.

Los métodos numéricos se dividen en dos categorías generales: métodos exactos y aproximados. Los primeros, como su nombre lo indica, buscan dar resultados exactos. No obstante, como están afectados por errores de redondeo, algunas veces dan resultados imprecisos. La magnitud del error de redondeo varía en cada sistema y depende de varios factores, como las dimensiones del sistema, su condición y el hecho de si la matriz de coeficientes es dispersa o densa. Además, la precisión de la computadora afectará el error de redondeo.

Se recomienda una estrategia de pivoteo en todo programa de computadora que realice métodos de eliminación exactos. Esa estrategia minimiza el error de redondeo y evita problemas como el de la división entre cero. Los algoritmos basados en la descomposición *LU* son los métodos que se eligen debido a su eficiencia y flexibilidad.

Tabla PT3.2 Comparación de las características de diversos métodos alternativos para encontrar soluciones de ecuaciones algebraicas lineales simultáneas.

Método	Estabilidad	Precisión	Rango de aplicación	Complejidad de programación	Comentarios
Gráfico	—	Pobre	Limitado	—	Puede tomar más tiempo que el método numérico pero puede ser útil para visualización
Regla de Cramer	—	Afectada por errores de redondeo	Limitado	—	Excesiva complejidad de cálculo para más de tres ecuaciones
Eliminación de Gauss (con pivoteo parcial)	—	Afectada por errores de redondeo	General	Moderada	
Descomposición *LU*	—	Afectada por errores de redondeo	General	Moderada	Método de eliminación preferido; permite el cálculo de la matriz inversa
Gauss-Seidel	Puede no converger si no es diagonalmente dominante	Excelente	Apropiada sólo para sistemas diagonalmente dominantes	Fácil	

Aunque los métodos de eliminación tienen gran utilidad, el uso de toda la matriz de los coeficientes puede ser limitante cuando se trate con sistemas dispersos muy grandes. Esto se debe a que gran parte de la memoria de la computadora se dedicaría a guardar ceros que no tienen significado. Para sistemas a bandas, hay técnicas para realizar métodos de eliminación sin tener que guardar todos los coeficientes de la matriz.

La técnica aproximada descrita en este libro se conoce como método de Gauss-Seidel, el cual difiere de las técnicas exactas porque emplea un esquema iterativo para obtener, progresivamente, estimaciones más cercanas a la solución. El efecto del error de redondeo es un punto discutible en el método de Gauss-Seidel, ya que se pueden continuar las iteraciones hasta que se obtenga la precisión deseada. Además, se pueden desarrollar versiones del método de Gauss-Seidel para utilizar de manera eficiente los requerimientos de almacenaje en computadora con sistemas dispersos. En

consecuencia, la técnica de Gauss-Seidel es útil para grandes sistemas de ecuaciones, donde los requerimientos de almacenaje podrían llevar a problemas significativos con las técnicas exactas.

La desventaja del método de Gauss-Seidel es que no siempre converge o algunas veces converge de manera lenta a la solución verdadera. Es confiable sólo para aquellos sistemas que son diagonalmente dominantes. Sin embargo, hay métodos de relajación que algunas veces contrarrestan tales desventajas. Además, como muchos sistemas de ecuaciones algebraicas lineales surgen de sistemas físicos que presentan dominancia diagonal, el método de Gauss-Seidel tiene gran utilidad para resolver problemas de ingeniería.

En resumen, varios factores serán relevantes en la elección de una técnica para un problema en particular que involucre ecuaciones algebraicas lineales. No obstante, como se mencionó antes, el tamaño y la densidad del sistema son factores particularmente importantes en la determinación de su elección.

PT3.5 RELACIONES Y FÓRMULAS IMPORTANTES

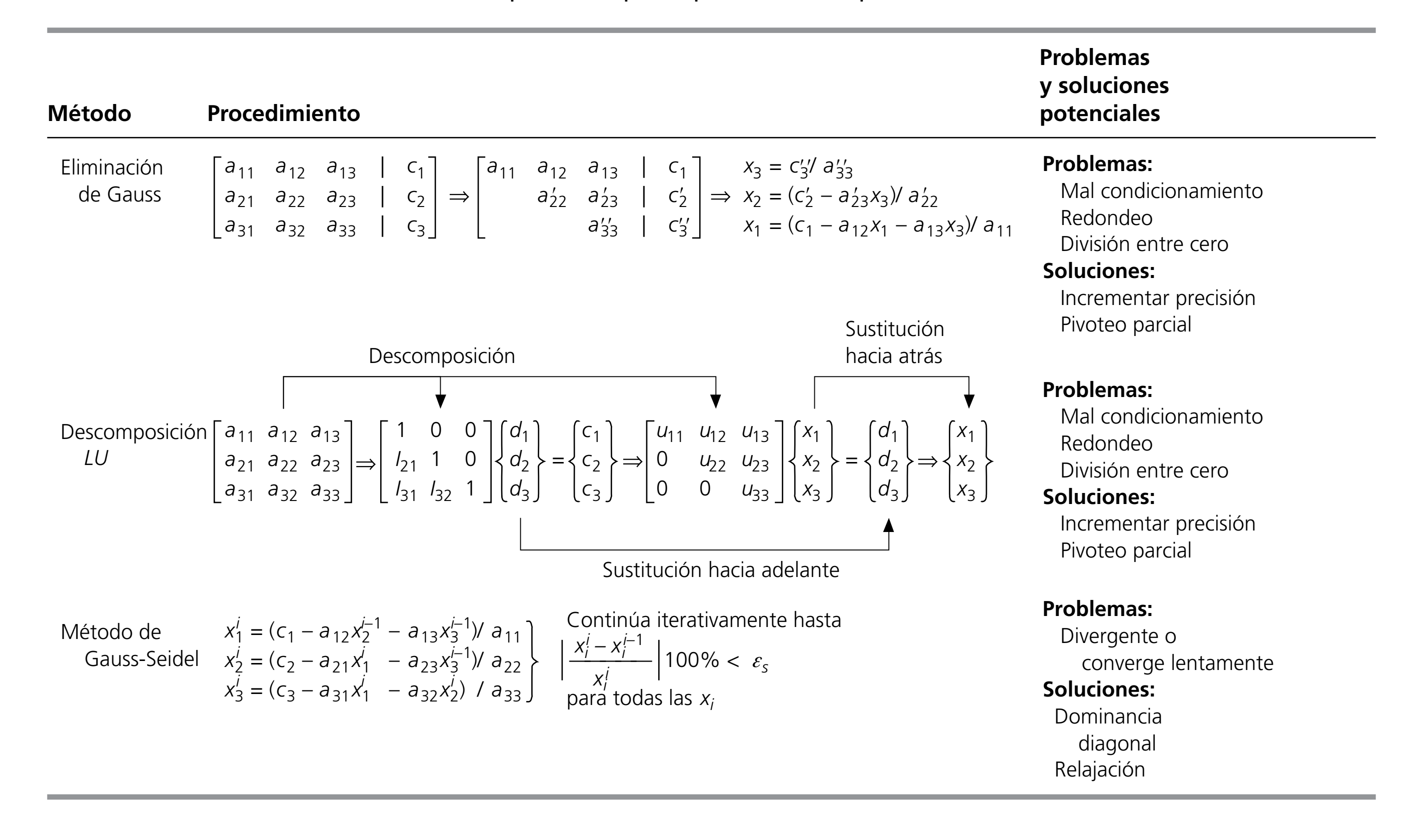
Cada una de las partes de este libro incluye una sección que resume fórmulas importantes. Aunque la parte tres no trata en realidad sólo con fórmulas, la tabla PT3.3 se emplea para resumir los algoritmos expuestos. La tabla proporciona una visión general, que será de gran ayuda para revisar y aclarar las principales diferencias entre los métodos.

PT3.6 MÉTODOS AVANZADOS Y REFERENCIAS ADICIONALES

Se pueden encontrar referencias generales acerca de la solución de ecuaciones lineales simultáneas en Faddeev y Faddeeva (1963), Stewart (1973), Varga (1962) y Young (1971). Ralston y Rabinowitz (1978) proporcionan un resumen general.

Hay muchas técnicas avanzadas para aumentar el ahorro de tiempo y/o espacio en la solución de ecuaciones algebraicas lineales. La mayoría de éstas se enfocan al aprovechamiento de las pro-

Tabla PT3.3 Resumen de información importante que se presenta en la parte tres.

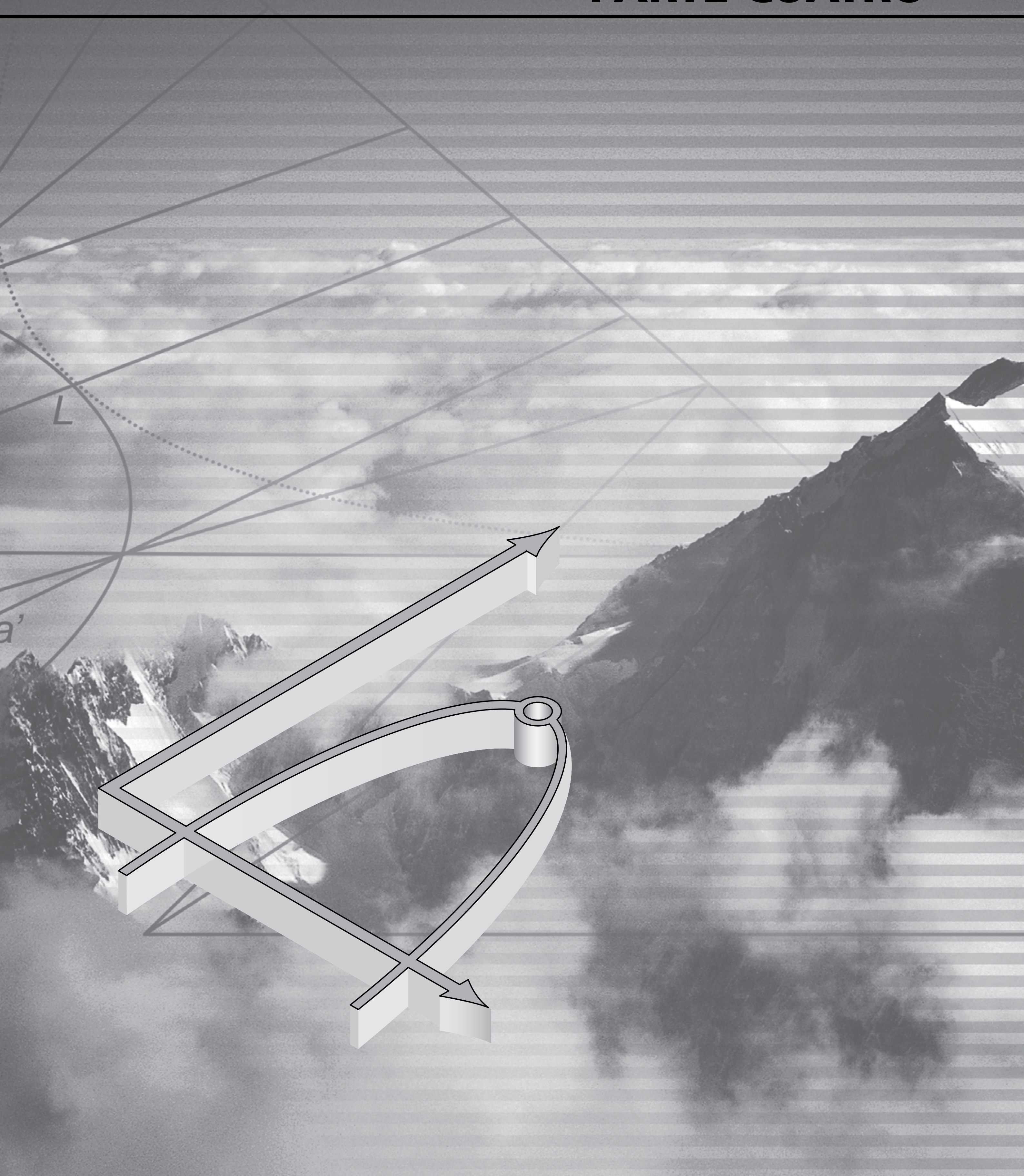


Método	Procedimiento	Problemas y soluciones potenciales
Eliminación de Gauss	$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \vert & c_1 \\ a_{21} & a_{22} & a_{23} & \vert & c_2 \\ a_{31} & a_{32} & a_{33} & \vert & c_3 \end{bmatrix} \Rightarrow \begin{bmatrix} a_{11} & a_{12} & a_{13} & \vert & c_1 \\ & a'_{22} & a'_{23} & \vert & c'_2 \\ & & a''_{33} & \vert & c''_3 \end{bmatrix} \Rightarrow \begin{matrix} x_3 = c''_3/a''_{33} \\ x_2 = (c'_2 - a'_{23}x_3)/a'_{22} \\ x_1 = (c_1 - a_{12}x_1 - a_{13}x_3)/a_{11} \end{matrix}$	**Problemas:** Mal condicionamiento; Redondeo; División entre cero. **Soluciones:** Incrementar precisión; Pivoteo parcial
Descomposición *LU*	Descomposición; Sustitución hacia atrás; Sustitución hacia adelante: $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \begin{Bmatrix} d_1 \\ d_2 \\ d_3 \end{Bmatrix} = \begin{Bmatrix} c_1 \\ c_2 \\ c_3 \end{Bmatrix} \Rightarrow \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} d_1 \\ d_2 \\ d_3 \end{Bmatrix} \Rightarrow \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix}$	**Problemas:** Mal condicionamiento; Redondeo; División entre cero. **Soluciones:** Incrementar precisión; Pivoteo parcial
Método de Gauss-Seidel	$\left.\begin{matrix} x_1^i = (c_1 - a_{12}x_2^{i-1} - a_{13}x_3^{i-1})/a_{11} \\ x_2^i = (c_2 - a_{21}x_1^i - a_{23}x_3^{i-1})/a_{22} \\ x_3^i = (c_3 - a_{31}x_1^i - a_{32}x_2^i)/a_{33} \end{matrix}\right\}$ Continúa iterativamente hasta $\left\vert \frac{x_i^i - x_i^{i-1}}{x_i^i} \right\vert 100\% < \varepsilon_s$ para todas las x_i	**Problemas:** Divergente o converge lentamente. **Soluciones:** Dominancia diagonal; Relajación

piedades de las ecuaciones, como simetría y a bandas. En particular se dispone de algoritmos que operan sobre matrices dispersas para convertirlas a un formato a bandas mínimo. Jacobs (1977) y Tewarson (1973) incluyen información sobre este tema. Una vez que se encuentran en un formato a bandas mínimo, existen diversas estrategias de solución eficientes: tal como el método de almacenamiento en una columna activa de Bathe y Wilson (1976).

Además de los conjuntos de ecuaciones $n \times n$, hay otros tipos de sistemas donde el número de ecuaciones, m, y el número de incógnitas, n, no son iguales. A los sistemas donde $m < n$ se les conoce como *subdeterminados*. En tales casos quizá no haya solución o tal vez haya más de una. Los sistemas donde $m > n$ se denominan *sobredeterminados*. En tales situaciones no hay, en general, solución exacta. Sin embargo, a menudo es posible desarrollar una solución que intente determinar soluciones que estén "lo más cercanas", para satisfacer todas las ecuaciones de manera simultánea. Un procedimiento común consiste en resolver la ecuación en un sentido de "mínimos cuadrados" (Lawson y Hanson, 1974; Wilkinson y Reinsch, 1971). Alternativamente, se pueden utilizar métodos de programación lineal, con los cuales las ecuaciones se resuelven en un sentido "óptimo", minimizando alguna función objetivo (Dantzig, 1963; Luenberger, 1984, y Rabinowitz, 1968). En el capítulo 15 se describe con mayor detalle este procedimiento.

PARTE CUATRO



OPTIMIZACIÓN

PT4.1 MOTIVACIÓN

La localización de raíces (parte dos) y la optimización están relacionadas, en el sentido de que ambas involucran valores iniciales y la búsqueda de un punto en una función. La diferencia fundamental entre ambos tipos de problemas se ilustra en la figura PT4.1. La localización de raíces es la búsqueda de los ceros de una función o funciones. En cambio, la *optimización* es la búsqueda ya sea del mínimo o del máximo.

El óptimo es el punto donde la curva es plana. En términos matemáticos, esto corresponde al valor de x donde la derivada $f'(x)$ es igual a cero. Además, la segunda derivada, $f''(x)$, indica si el óptimo es un mínimo o un máximo: si $f''(x) < 0$, el punto es un máximo; si $f''(x) > 0$, el punto es un mínimo.

Si comprendemos ahora la relación entre las raíces y el óptimo, es posible sugerir una estrategia para determinar este último; es decir, se puede derivar a la función y localizar la raíz (el cero) de la nueva función. De hecho, algunos métodos de optimización tratan de encontrar un óptimo resolviendo el problema de encontrar la raíz: $f'(x) = 0$. Deberá observarse que tales búsquedas con frecuencia se complican porque $f'(x)$ no se puede obtener analíticamente. Por lo tanto, es necesario usar aproximaciones por diferencias finitas para estimar la derivada.

Más allá de ver la optimización como un problema de raíces, deberá observarse que la tarea de localizar el óptimo está reforzada por una estructura matemática extra que no es parte del encontrar una simple raíz. Esto tiende a hacer de la optimización una tarea más manejable, en particular con casos multidimensionales.

PT4.1.1 Métodos sin computadora e historia

Como se mencionó antes, los métodos de cálculo diferencial aún se utilizan para determinar soluciones óptimas. Todos los estudiantes de ciencias e ingeniería recuerdan haber resuelto problemas de máximos y mínimos mediante la determinación de las primeras derivadas de las funciones en sus cursos sobre cálculo. Bernoulli, Euler, Lagrange y otros establecieron los fundamentos del cálculo de variaciones, el cual trata con la minimización de funciones. El método de los multiplicadores de Lagrange se desarrolló para optimizar problemas con restricciones, es decir, problemas de optimización donde las variables están limitadas en alguna forma.

El primer avance de importancia en los procedimientos numéricos ocurrió con el desarrollo de las computadoras digitales después de la Segunda Guerra Mundial. Koopmans, en el Reino Unido, y Kantorovich, en la ex Unión Soviética, trabajaron en forma independiente sobre el problema general de distribución a bajo costo de artículos y productos. En 1947, un alumno de Koopman, Dantzig, inventó el *método simplex* para resolver problemas de programación lineal. Este método abrió el camino a muchos investigadores hacia otros métodos de optimización con restricciones; entre los más notables se encuentran Charnes y sus colegas. Los métodos de optimización con restricciones también se desarrollaron en forma rápida debido a la disponibilidad tan amplia de computadoras.

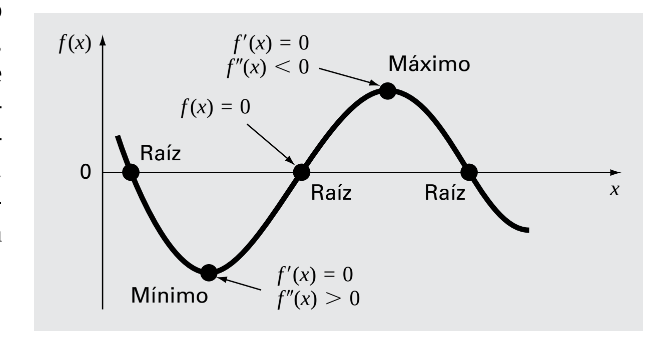


Figura PT4.1
Una función de una sola variable ilustra la diferencia entre las raíces y el óptimo.

PT4.1.2 Optimización y la práctica en ingeniería

La mayoría de los modelos matemáticos con que hemos tratado hasta ahora han sido *descriptivos*. Es decir, se han obtenido para simular el comportamiento de un dispositivo o sistema en inge-

nería. En cambio, la optimización tiene que ver con la determinación del "mejor resultado", o solución óptima, de un problema. Así, en el contexto del modelado, se les llama con frecuencia modelos *prescriptivos*, puesto que sirven para señalar un curso de acción o el mejor diseño.

Los ingenieros continuamente tienen que diseñar dispositivos y productos que realicen tareas de manera eficiente. Al hacerlo de esta manera, están restringidos por las limitaciones del mundo físico. Además, deben mantener costos bajos. Así, los ingenieros siempre se enfrentan a problemas de optimización que equilibren el funcionamiento y las limitaciones. Algunos ejemplos comunes se mencionan en la tabla PT4.1. El siguiente ejemplo fue desarrollado para ayudarlo a obtener una visión de la manera en que se pueden formular tales problemas.

EJEMPLO PT.4.1 Optimización del costo de un paracaídas

Planteamiento del problema A lo largo de este libro, hemos utilizado la caída de un paracaidista para ilustrar diversos temas básicos para la solución de problemas con métodos numéricos. Usted puede haber notado que ninguno de tales ejemplos se ocupó de lo que pasa después de que el paracaídas se abre. En este ejemplo examinaremos un caso donde el paracaídas se abre, y nos interesa predecir la velocidad de impacto con el suelo.

Usted es un ingeniero que trabaja para una institución que lleva abastecimientos a los refugiados en una zona de guerra. Los abastecimientos se arrojarán a baja altitud (500 m), de tal forma que la caída no sea detectada y que los abastecimientos caigan tan cerca como sea posible del campo de refugiados. Los paracaídas se abren en forma inmediata al salir del aeroplano. Para reducir daños, la velocidad vertical de impacto debe ser menor a un valor crítico $v_c = 20$ m/s.

El paracaídas que se usa para la caída se ilustra en la figura PT4.2. El área de la sección transversal del paracaídas es la de una semiesfera,

$$A = 2\pi r^2 \tag{PT4.1}$$

La longitud de cada una de las 16 cuerdas, que unen al paracaídas con la masa, está relacionada con el radio del paracaídas mediante

$$\ell = \sqrt{2}r \tag{PT4.2}$$

Usted sabe que la fuerza de arrastre del paracaídas es una función lineal del área de su sección transversal descrita con la siguiente fórmula:

$$c = k_c A \tag{PT4.3}$$

donde c = coeficiente de arrastre (kg/s) y k_c = una constante de proporcionalidad que parametriza el efecto del área sobre el arrastre [kg/(s · m^2)].

Tabla PT4.1 Algunos ejemplos comunes de problemas de optimización en ingeniería.

- Diseño de un avión con peso mínimo y resistencia máxima.
- Trayectorias óptimas de vehículos espaciales.
- Diseño de estructuras en la ingeniería civil con un mínimo costo.
- Planeación de obras para el abastecimiento de agua, como presas, que permitan disminuir daños por inundación, mientras se obtiene máxima potencia hidráulica.
- Predicción del comportamiento estructural minimizando la energía potencial.
- Determinación del corte de materiales con un mínimo costo.
- Diseño de bombas y equipos de transferencia de calor con una máxima eficiencia.
- Maximización de la potencia de salida de circuitos eléctricos y de maquinaria, mientras se minimiza la generación de calor.
- Ruta más corta de un vendedor que recorre varias ciudades durante un viaje de negocios.
- Planeación y programación óptimas.
- Análisis estadístico y modelado con un mínimo error.
- Redes de tubería óptimas.
- Control de inventario.
- Planeación del mantenimiento para minimizar costos.
- Minimización de tiempos de espera.
- Diseño de sistemas de tratamiento de residuos para cumplir con estándares de calidad del agua a bajo costo.

También, se puede dividir la carga completa en tantos paquetes como se quiera. Es decir, la masa de cada paquete se calcula así

$$m = \frac{M_t}{n}$$

donde m = masa de cada paquete (kg), M_t = carga total que habrá de arrojarse (kg) y n = número total de paquetes.

Por último, el costo de cada paracaídas está relacionado con su tamaño en una forma no lineal,

$$\text{Costo por paracaídas} = c_0 + c_1\ell + c_2A^2 \tag{PT4.4}$$

donde c_0, c_1 y c_2 son coeficientes de costo. El término constante, c_0, es el costo base de los paracaídas. La relación no lineal entre costo y área se debe a que la fabricación de los paracaídas de gran tamaño es más complicada que la de los paracaídas pequeños.

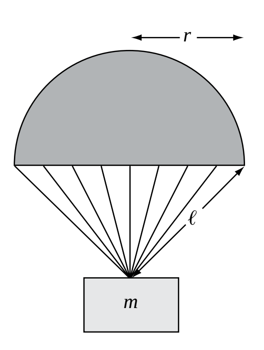


Figura PT4.2
Un paracaídas abierto.

Determine el tamaño (r) y el número de paracaídas (n) que se obtienen a un mínimo costo y que, al mismo tiempo, satisfacen el requerimiento de lograr una velocidad de impacto suficientemente pequeña.

Solución El objetivo aquí consiste en determinar la cantidad y el tamaño de los paracaídas que minimicen el costo de la operación. El problema tiene restricciones, ya que los paquetes deben tener una velocidad de impacto menor al valor crítico.

El costo se calcula al multiplicar el valor de un solo paracaídas [ecuación (PT4.4)] por el número de paracaídas (n). Así, la función que usted debe minimizar, llamada formalmente *función objetivo*, se escribe como

$$\text{Minimizar } C = n(c_0 + c_1\ell + c_2A^2) \tag{PT4.5}$$

donde C = costo ($) y A y ℓ se calculan con las ecuaciones (PT4.1) y (PT4.2), respectivamente.

A continuación, se deben especificar las *restricciones*. En este problema existen dos restricciones. Primera, la velocidad de impacto debe ser igual o menor que la velocidad crítica,

$$v \le v_c \tag{PT4.6}$$

Segunda, el número de paquetes debe ser un entero mayor o igual a 1,

$$n \ge 1 \tag{PT4.7}$$

donde n es un entero.

En este momento, ya se ha formulado el problema de optimización. Como se observa, es un problema con restricciones no lineal.

Aunque el problema se ha formulado completamente, se debe tener en cuenta algo más: ¿cómo se determina la velocidad de impacto v? Recuerde del capítulo 1 que la velocidad de un objeto que cae se calcula así:

$$v = \frac{gm}{c}(1 - e^{-(c/m)t}) \tag{1.10}$$

donde v = velocidad (m/s), g = aceleración de la gravedad (m/s^2), m = masa (kg) y t = tiempo (s).

Aunque la ecuación (1.10) proporciona una relación entre v y t, lo que se necesita saber es en cuánto tiempo cae la masa. Por lo tanto, es necesaria una relación entre la distancia de caída z y el tiempo de caída t. La distancia de caída se calcula a partir de la velocidad en la ecuación (1.10) mediante la integración

$$z = \int_0^t \frac{gm}{c}(1 - e^{-(c/m)t})\, dt \tag{PT4.8}$$

Esta integral se evalúa para obtener

$$z = z_0 - \frac{gm}{c}t + \frac{gm^2}{c^2}(1 - e^{-(c/m)t}) \tag{PT4.9}$$

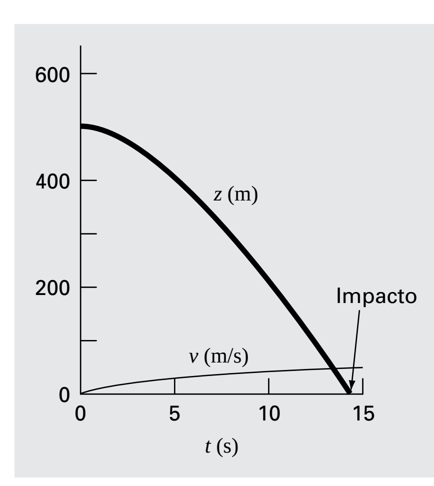


Figura PT4.3
La altura z y la velocidad v de un paracaídas abierto conforme cae al suelo (z = 0).

donde z_0 = altura inicial (m). Esta función, como muestra la gráfica de la figura PT4.3, ofrece una manera de predecir z conociendo t.

Sin embargo, no se necesita z como función de t para resolver este problema. Lo que necesitamos es el tiempo requerido por el paquete, al caer, la distancia z_0. Así, se reconoce que tenemos que reformular la ecuación (PT4.9) como un problema de determinación de raíces. Esto es, se debe encontrar el tiempo en el que z toma el valor de cero,

$$f(t) = 0 = z_0 - \frac{gm}{c}t + \frac{gm^2}{c^2}(1 - e^{-(c/m)t}) \tag{PT4.10}$$

Una vez que se calcula el tiempo de impacto, se sustituye en la ecuación (1.10) con la finalidad de obtener la velocidad de impacto.

El planteamiento del problema sería entonces

$$\text{Minimizar } C = n(c_0 + c_1\ell + c_2A^2) \tag{PT4.11}$$

sujeta a

$$v \leq v_c \tag{PT4.12}$$

$$n \geq 1 \tag{PT4.13}$$

donde

$$A = 2\pi r^2 \tag{PT4.14}$$

$$\ell = \sqrt{2}r \tag{PT4.15}$$

$$c = k_cA \tag{PT4.16}$$

$$m = \frac{M_t}{n} \tag{PT4.17}$$

$$t = \text{raíz}\left[z_0 - \frac{gm}{c}t + \frac{gm^2}{c^2}(1 - e^{-(c/m)t})\right] \tag{PT4.18}$$

$$v = \frac{gm}{c}(1 - e^{-(c/m)t}) \tag{PT4.19}$$

Resolveremos este problema en el ejemplo 15.4 al final del capítulo 15. Por ahora reconozca que este problema tiene la mayoría de los elementos fundamentales de otros problemas de optimización, que usted enfrentará en la práctica de la ingeniería. Éstos son

- El problema involucrará una *función objetivo* que se optimizará.
- Tendrá también un número de *variables de diseño*. Éstas pueden ser números reales o enteros. En nuestro ejemplo, dichas variables son r (real) y n (entero).
- El problema incluye *restricciones* que consideran las limitaciones bajo las cuales se trabaja.

Plantearemos una reflexión más antes de proceder. Aunque la función objetivo y las restricciones quizá, en forma superficial, parezcan ecuaciones simples [por ejemplo, la ecuación (PT4.12)], de hecho, pueden ser sólo la "punta del *iceberg*". Es decir, pueden basarse en modelos y dependencias complicadas. Por ejemplo, como en este caso, llegan a involucrar otros métodos numéricos [ecuación (PT4.18)], lo cual significa que las relaciones funcionales que usted estará usando podrían representar cálculos largos y complicados. Por lo que las técnicas que permitan encontrar la solución óptima, y que al mismo tiempo simplifiquen las evaluaciones de las funciones, serán valiosas en extremo.

PT4.2 ANTECEDENTES MATEMÁTICOS

Existen bastantes conceptos matemáticos que son la base de la optimización. Como creemos que para usted éstos serán más relevantes en su forma contextual, se dejará el análisis de los prerrequisitos matemáticos específicos hasta que se ocupen. Por ejemplo, se analizarán los importantes conceptos

del gradiente y el hessiano al inicio del capítulo 14, que trata sobre optimización multivariable sin restricciones. Mientras tanto, ahora nos limitaremos al tema más general de cómo se clasifican los problemas de optimización.

Un problema de *programación matemática* u *optimización* generalmente se puede establecer como

Determine $\boldsymbol{x}$, que minimiza o maximiza $f(\boldsymbol{x})$

sujeto a

$$d_i(\boldsymbol{x}) \leq a_i \qquad i = 1, 2, \ldots, m \tag{PT4.20}$$

$$e_i(\boldsymbol{x}) = b_i \qquad i = 1, 2, \ldots, p \tag{PT4.21}$$

donde $\boldsymbol{x}$ es un *vector de diseño n*-dimensional; $f(\boldsymbol{x})$ *es la función objetivo*; $d_i(\boldsymbol{x})$ son las *restricciones de desigualdad*; $e_i(\boldsymbol{x})$ son las *restricciones de igualdad*, y a_i y b_i son constantes.

Los problemas de optimización se clasifican considerando la forma de $f(\boldsymbol{x})$:

- Si $f(\boldsymbol{x})$ y las restricciones son lineales, tenemos un problema de *programación lineal*.
- Si $f(\boldsymbol{x})$ es cuadrática y las restricciones son lineales, tenemos un problema de *programación cuadrática*.
- Si $f(\boldsymbol{x})$ no es lineal ni cuadrática y/o las restricciones no son lineales, tenemos un problema de *programación no lineal*.

Se dice también que, cuando las ecuaciones (PT4.20) y (PT4.21) se incluyen, se tiene un problema de *optimización con restricciones*; de otra forma, se trata de un problema de *optimización sin restricciones.*

Observe que en problemas con restricciones, los grados de libertad están dados por n-p-m. Generalmente, para obtener una solución, $p + m$ debe ser $\leq n$. Si $p + m > n$, se dice que el problema está *sobrerrestringido.*

Otra forma de clasificar los problemas de optimización es según su dimensionalidad. En general se dividen en *unidimensionales* y *multidimensionales.* Como su nombre lo indica, los primeros involucran funciones que dependen de una sola variable independiente. Como en la figura PT4.4*a*, la búsqueda consiste, entonces, en ascender o descender picos y valles unidimensionales. Los *problemas multidimensionales* implican funciones que dependen de dos o más variables independientes. En el mismo sentido, la optimización bidimensional, de nuevo, se visualiza como una búsqueda de picos y valles (figura PT4.4*b*). Sin embargo, justo como en un paseo campestre, no estamos limitados a caminar en una sola dirección; en lugar de esto se examina la *topografía* para alcanzar el objetivo en forma eficiente.

Finalmente, el proceso de encontrar un máximo o de encontrar un mínimo es, en esencia, idéntico, ya que un mismo valor, por ejemplo $\boldsymbol{x}^*$, minimiza $f(\boldsymbol{x})$ y maximiza $-f(\boldsymbol{x})$. Esta equivalencia se ilustra en forma gráfica, para una función unidimensional, en la figura PT4.4*a*.

PT4.3 ORIENTACIÓN

Resulta útil alguna orientación antes de desarrollar los métodos numéricos para la optimización. Lo siguiente lleva la intención de dar una visión general del material en la parte cuatro. Además, se presentan algunos objetivos para ayudarlo a enfocar sus esfuerzos cuando se estudie el material.

PT4.3.1 Alcance y presentación preliminar

La figura PT4.5 es una representación esquemática de la organización de la parte cuatro. Examine esta figura con cuidado, comenzando desde arriba y después yendo en sentido de las manecillas del reloj.

Después de la presente introducción, el *capítulo 13* se dedica a la *optimización unidimensional sin restricciones.* Se presentan métodos para determinar el mínimo o el máximo de una función con una sola variable. Se examinan tres métodos: *búsqueda de la sección dorada, interpolación cuadrática* y el *método de Newton*. También se describe un enfoque híbrido avanzado, el *método de Brent*, que combina la confiabilidad de la sección dorada con la rapidez de la interpolación cuadrática.



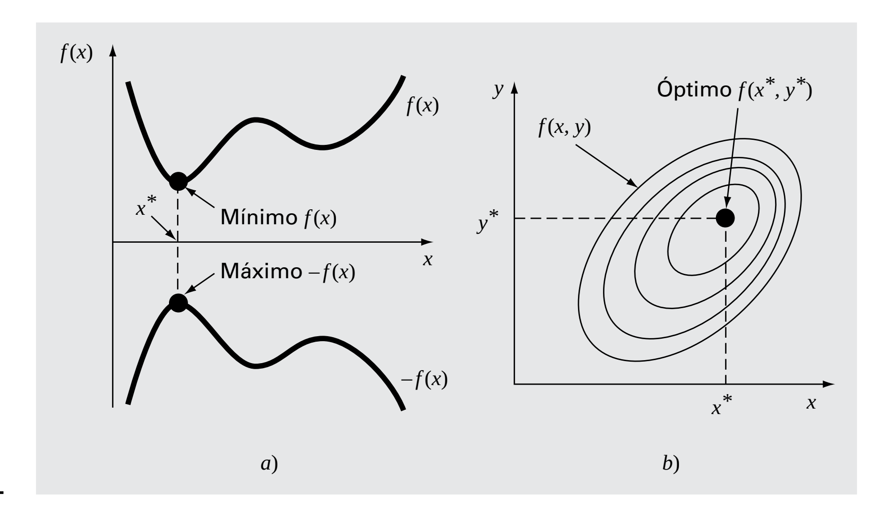


Figura PT4.4
a) Optimización unidimensional. Esta figura también ilustra cómo la minimización de $f(x)$ es equivalente a la maximización de $-f(x)$. *b*) Optimización bidimensional. Observe que esta figura puede tomarse para representar ya sea una maximización (los contornos aumentan de elevación hasta un máximo como en una montaña), o una minimización (los contornos disminuyen de elevación hasta un mínimo como un valle).

El *capítulo 14* cubre dos tipos generales de métodos para resolver problemas de *optimización multidimensional sin restricciones*. Los *métodos directos*, como *búsquedas aleatorias*, *búsquedas univariables* y *búsquedas de patrones*, no requieren la evaluación de las derivadas de la función. Por otro lado, los *métodos de gradiente* utilizan la primera o la segunda derivada para encontrar el óptimo. En este capítulo se introduce el *gradiente* y el *hessiano*, que son las representaciones multidimensionales de la primera y la segunda derivada. El método de *paso ascendente/descendente* se estudia después con detalle. A esto le siguen descripciones de algunos métodos avanzados: el *gradiente conjugado*, el *método de Newton*, el *método de Marquardt* y los *métodos cuasi-Newton*.

El *capítulo 15* se dedica a la *optimización con restricciones*. La *programación lineal* se describe con detalle usando tanto la representación gráfica como el *método simplex*. El análisis detallado de *optimización con restricciones no lineal* está fuera del alcance de este texto; no obstante, se ofrece una visión general de los principales métodos. Además, se ilustra cómo tales problemas (junto con los estudiados en los capítulos 13 y 14) se resuelven con paquetes de software, como Excel, MATLAB y Mathcad.

En el *capítulo 16* se extienden los conceptos anteriores a problemas que se presentan en la ingeniería. Se utilizan las aplicaciones en ingeniería para ilustrar cómo se formulan los problemas de optimización, y para dar una visión sobre la aplicación de las técnicas de solución en la práctica profesional.

Se incluye un epílogo al final de la parte cuatro. Éste contiene un repaso de los métodos analizados en los capítulos 13, 14 y 15. Dicho repaso da una descripción de las ventajas y desventajas relacionadas con el uso apropiado de cada técnica. Esta sección también presenta referencias acerca de algunos métodos numéricos que van más allá del alcance de este libro.

PT4.3.2 Metas y objetivos

Objetivos de estudio Después de estudiar la parte cuatro, usted tendrá suficiente información para abordar con éxito una amplia variedad de problemas que se presentan en la ingeniería, relacionados con la optimización. En general, usted deberá dominar las técnicas, habrá aprendido a evaluar su confiabilidad y será capaz de analizar métodos alternativos para un problema específico. Además, de estas metas generales, deberán asimilarse los conceptos específicos dados en la tabla PT4.2 para un aprendizaje completo del material de la parte cuatro.

Objetivos de cómputo Usted deberá ser capaz de escribir un subprograma que lleve a cabo una búsqueda simple unidimensional (como la búsqueda de la sección dorada o la interpolación cuadrática) y multidimensional (como el método de búsqueda aleatoria). Además, los paquetes de software como Excel, MATLAB o Mathcad tienen varias capacidades para optimización. Usted puede usar esta parte del libro para familiarizarse con todas estas capacidades.

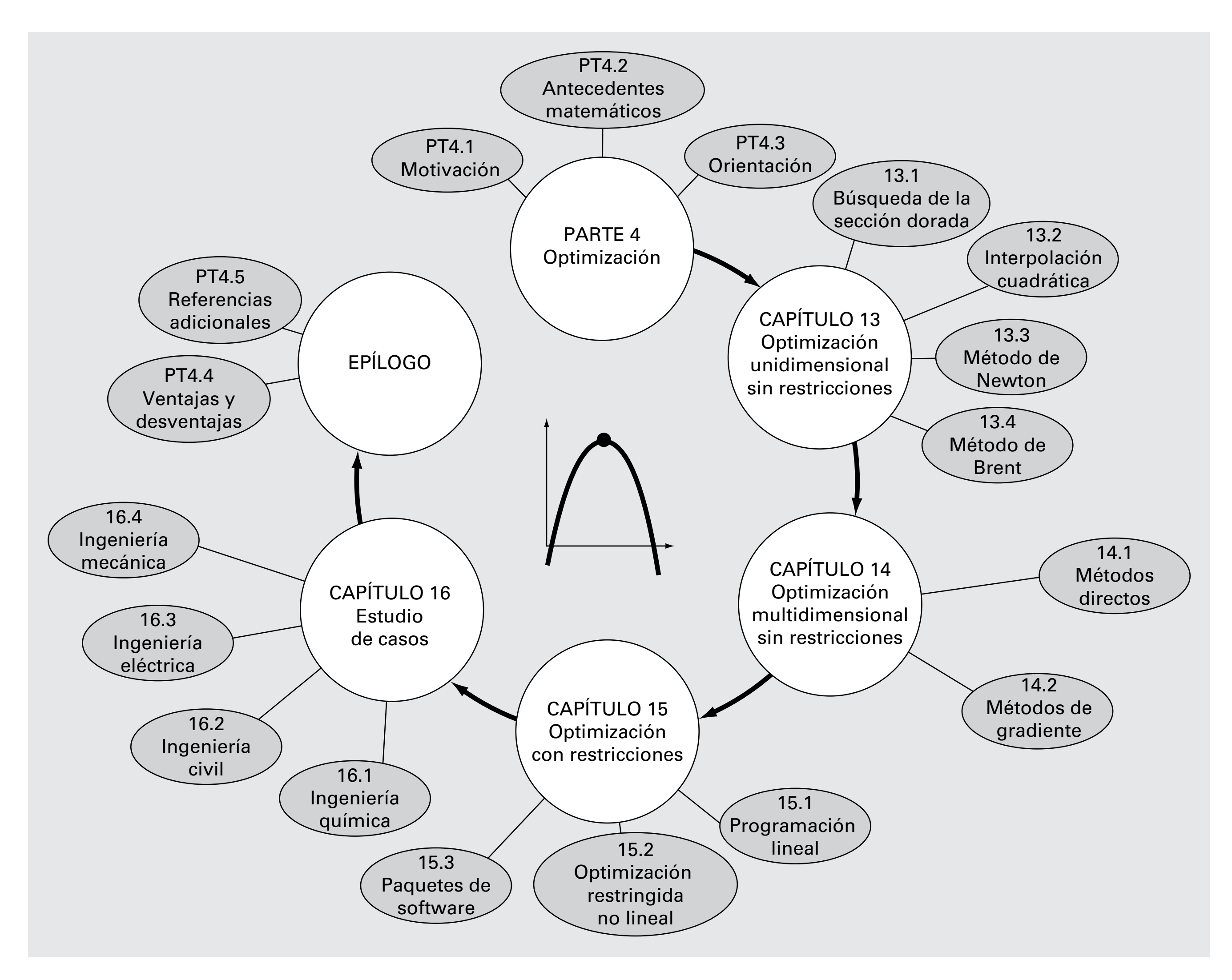


Figura PT4.5
Representación de la organización del material en la parte cuatro: Optimización.

Tabla PT4.2 Objetivos específicos de estudio de la parte cuatro.

1. Entender por qué y dónde se presenta la optimización al resolver problemas de ingeniería.
2. Comprender los principales elementos del problema de optimización general: función objetivo, variables de decisión y restricciones.
3. Ser capaz de distinguir entre la optimización lineal y la no lineal, y entre problemas con restricciones y sin restricciones.
4. Poder definir la razón dorada y comprender cómo hace que la optimización unidimensional sea eficiente.
5. Localizar el óptimo de una función en una sola variable mediante la búsqueda de la sección dorada, la interpolación cuadrática y el método de Newton. También, reconocer las ventajas y desventajas de tales métodos, especialmente en relación con los valores iniciales y la convergencia.
6. Entender cómo el método de optimización de Brent combina la confiabilidad de la búsqueda de la sección dorada con la rapidez de la interpolación parabólica.
7. Escribir un programa y encontrar el óptimo de una función multivariada usando la búsqueda aleatoria.
8. Comprender las ideas de los patrones de búsqueda, las direcciones conjugadas y el método de Powell.
9. Definir y evaluar el gradiente y el hessiano de una función multivariada, tanto en forma analítica como numérica.
10. Calcular a mano el óptimo de una función con dos variables, usando el método de paso ascendente-descendente.
11. Comprender las ideas básicas de los métodos del gradiente conjugado, de Newton, de Marquardt y de cuasi-Newton. En particular, entender las ventajas y las desventajas de los diferentes métodos, y reconocer cómo cada uno mejora el de paso ascendente-descendente.
12. Reconocer y plantear un problema de programación lineal para representar problemas aplicables a la ingeniería.
13. Resolver un problema de programación lineal bidimensional con ambos métodos: el gráfico y el simplex.
14. Comprender los cuatro posibles resultados de un problema de programación lineal.
15. Plantear y resolver problemas de optimización con restricciones no lineales utilizando un paquete de software.

CAPÍTULO 13

Optimización unidimensional sin restricciones

Esta sección describirá técnicas para encontrar el mínimo o el máximo de una función de una sola variable, $f(x)$. Una imagen útil que muestra lo anterior es la consideración unidimensional a la "montaña rusa", como la función representada en la figura 13.1. Recuerde que en la parte dos, la localización de una raíz fue complicada por el hecho de que una sola función puede tener varias raíces. De manera similar, los valores óptimos tanto locales como globales pueden presentarse en problemas de optimización. A tales casos se les llama *multimodales*. En casi todos los ejemplos, estaremos interesados en encontrar el valor máximo o mínimo absoluto de una función. Así, debemos cuidar de no confundir un óptimo local con un óptimo global.

Distinguir un extremo global de un extremo local puede ser generalmente un problema difícil. Existen tres formas comunes de resolver este problema. Primero, una idea del comportamiento de las funciones unidimensionales algunas veces llega a obtenerse en forma gráfica. Segundo, determinar el valor óptimo con base en valores iniciales, los cuales varían ampliamente y son generados quizá en forma aleatoria, para después seleccionar el mayor de éstos como el global. Por último, cambiar el punto de inicio asociado con un óptimo local y observar si la rutina empleada da un mejor punto, o siempre regresa al mismo punto. Aunque estos métodos tienen su utilidad, el hecho es que en algunos problemas (usualmente los más grandes) no existe una forma práctica de asegurarse de que se ha localizado un valor óptimo global. Sin embargo, aunque debe tenerse cuidado se tiene la fortuna de que en muchos problemas de la ingeniería se localiza el óptimo global en forma no ambigua.

Como en la localización de raíces, los problemas de optimización unidimensionales se pueden dividir en métodos cerrados y métodos abiertos. Como se describirá en la próxima sección, la búsqueda de la sección dorada es un ejemplo de un método cerrado que depende de los valores iniciales que encierran un solo valor óptimo. Esto va seguido de un enfoque alternativo, la interpolación parabólica, que a menudo converge más rápidamente que la búsqueda de la sección dorada, pero algunas veces diverge.

Otro método que se describe en este capítulo es un método abierto basado en la idea, proveniente del cálculo, de que se pueden hallar el mínimo o el máximo resolviendo $f'(x) = 0$. Esto reduce el problema de optimización a encontrar la raíz de $f'(x)$ usando técnicas del tipo que se describen en la parte dos. Se demostrará una versión de este enfoque: el método de Newton.

Finalmente se describe un enfoque híbrido avanzado: el *método de Brent*. Éste combina la confiabilidad de la búsqueda de la sección dorada con la rapidez de la interpolación parabólica.

Figura 13.1
Una función que se aproxima asintóticamente a cero en más y menos ∞ y que tiene dos puntos máximos y dos puntos mínimos en la vecindad del origen. Los dos puntos a la derecha son los óptimos locales; mientras que los dos de la izquierda son globales.

13.1 BÚSQUEDA DE LA SECCIÓN DORADA

En la búsqueda de la raíz de una ecuación no lineal, el objetivo era encontrar el valor de x que diera *cero* al sustituir en la función $f(x)$. La *optimización en una sola variable* tiene como objetivo encontrar el valor de x que da un *extremo*, ya sea un máximo o un mínimo de $f(x)$.

La búsqueda de la sección dorada es una técnica de búsqueda para una sola variable, sencilla y de propósito general. Es igual en esencia al método de la bisección para localizar raíces (capítulo 5). Recuerde que la bisección depende de la definición de un

intervalo, especificado por los valores iniciales inferior (x_l) y superior (x_u), que encierran una sola raíz. La presencia de una raíz entre estos límites se verificó determinando que $f(x_l)$ y $f(x_u)$ tuvieran signos diferentes. La raíz se estima entonces como el punto medio de este intervalo,

$$x_r = \frac{x_l + x_u}{2}$$

El paso final en una iteración por bisección permite determinar un intervalo más pequeño. Esto se logra al reemplazar cualquiera de los límites, x_l o x_u, que tuvieran un valor de la función con el mismo signo que $f(x_r)$. Un efecto útil de este método es que el nuevo valor x_r reemplazará a uno de los límites anteriores.

Es posible desarrollar un procedimiento similar para localizar el valor óptimo de una función unidimensional. Por simplicidad, nos concentraremos en el problema de encontrar un máximo. Cuando se analice el algoritmo de cómputo, se describirán las pequeñas modificaciones necesarias para determinar un mínimo.

Como en el método de la bisección, se puede comenzar por definir un intervalo que contenga una sola respuesta. Es decir, el intervalo deberá contener un solo máximo, y por esto se llama *unimodal*. Podemos adoptar la misma nomenclatura que para la bisección, donde x_l y x_u definen los límites inferior y superior, respectivamente, del intervalo. Sin embargo, a diferencia de la bisección se necesita una nueva estrategia para encontrar un máximo dentro del intervalo. En vez de usar solamente dos valores de la función (los cuales son suficientes para detectar un cambio de signo y, por lo tanto, un cero), se necesitarán tres valores de la función para detectar si hay un máximo. Así, hay que escoger un punto más dentro del intervalo. Después, hay que tomar un cuarto punto. La prueba para el máximo podrá aplicarse para determinar si el máximo se encuentra dentro de los primeros tres o de los últimos tres puntos.

La clave para hacer eficiente este procedimiento es la adecuada elección de los puntos intermedios. Como en la bisección, la meta es minimizar las evaluaciones de la función reemplazando los valores anteriores con los nuevos. Esta meta se puede alcanzar especificando que las siguientes dos condiciones se satisfagan (figura 13.2):

$$\ell_0 = \ell_1 + \ell_2 \tag{13.1}$$

$$\frac{\ell_1}{\ell_0} = \frac{\ell_2}{\ell_1} \tag{13.2}$$

La primera condición especifica que la suma de las dos sublongitudes ℓ_1 y ℓ_2 debe ser igual a la longitud original del intervalo. La segunda indica que el cociente o razón entre las longitudes debe ser igual. La ecuación (13.1) se sustituye en la (13.2),

$$\frac{\ell_1}{\ell_1 + \ell_2} = \frac{\ell_2}{\ell_1} \tag{13.3}$$

Si se toma el recíproco y $R = \ell_2/\ell_1$, se llega a

$$1 + R = \frac{1}{R} \tag{13.4}$$

o

$$R^2 + R - 1 = 0 \tag{13.5}$$

de la cual se obtiene la raíz positiva

$$R = \frac{-1 + \sqrt{1 - 4(-1)}}{2} = \frac{\sqrt{5} - 1}{2} = 0.61803\ldots \tag{13.6}$$

Este valor, que se conoce desde la antigüedad, se llama *razón dorada* o *razón áurea* (véase el cuadro 13.l). Como permite encontrar el valor óptimo en forma eficiente, es el elemento clave del método de la sección dorada que hemos estado desarrollando. Ahora construyamos un algoritmo para implementar este procedimiento en la computadora.

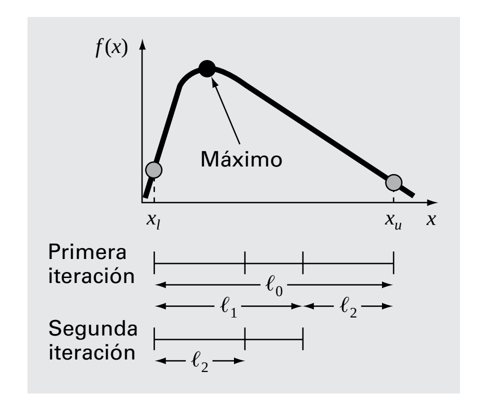


Figura 13.2
El paso inicial en el algoritmo de búsqueda de la sección dorada consiste en elegir dos puntos interiores de acuerdo con la razón dorada.

Cuadro 13.1 La razón dorada y los números de Fibonacci

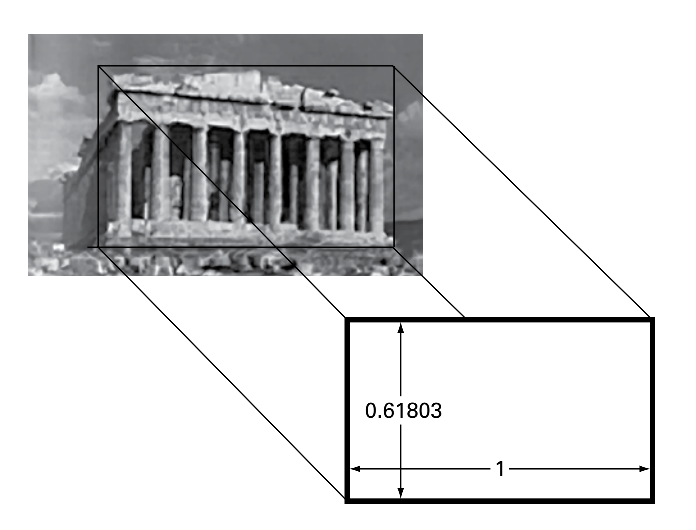


Figura 13.3
El Partenón de Atenas, Grecia, fue construido en el siglo V antes de Cristo. Sus dimensiones frontales se ajustan casi exactamente a un rectángulo dorado.

En muchas culturas, a ciertos números se les otorgan algunas cualidades. Por ejemplo, en Occidente se suele decir "el 7 de la suerte" y "el funesto viernes 13". Los antiguos griegos llamaron al siguiente número la "razón dorada" o áurea:

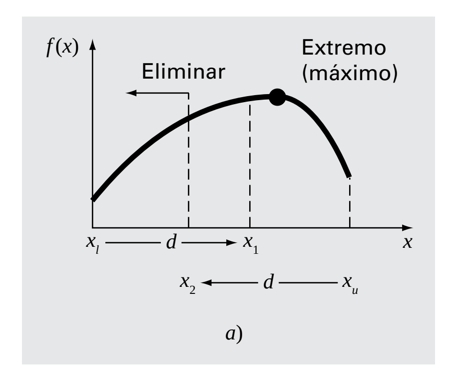
$$\frac{\sqrt{5}-1}{2} = 0.61803\ldots$$

Esta razón fue empleada con un gran número de propósitos, incluyendo el desarrollo del rectángulo de la figura 13.3. Tales proporciones fueron consideradas por los griegos como estéticamente agradables. Entre otras cosas, muchos de los templos siguieron esta forma.

La razón dorada se relaciona con una importante sucesión matemática conocida como los *números de Fibonacci*, que son

0, 1, 1, 2, 3, 5, 8, 13, 21, 34,...

Cada número después de los dos primeros representa la suma de los dos precedentes. Esta secuencia aparece en diversas áreas de la ciencia y la ingeniería. En el contexto del presente análisis, una interesante propiedad de la sucesión de Fibonacci relaciona la razón entre números consecutivos de la serie; es decir, 0/1 = 0, 1/1 = 1, 1/2 = 0.5, 2/3 = 0.667, 3/5 = 0.6, 5/8 = 0.625, 8/13 = 0.615, y así sucesivamente. La razón entre números consecutivos se va aproximando a la razón dorada.



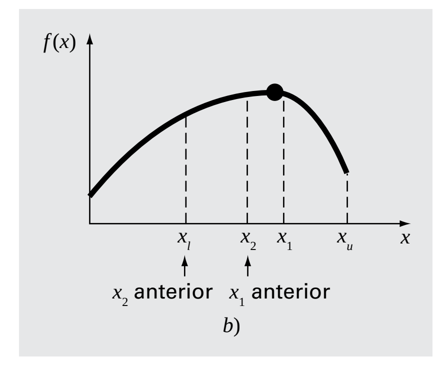


Figura 13.4
a) El paso inicial del algoritmo de búsqueda de la sección dorada involucra escoger dos puntos interiores de acuerdo con la razón dorada. *b*) El segundo paso implica definir un nuevo intervalo que incluya el valor óptimo.

Como se mencionó antes y se ilustra en la figura 13.4, el método comienza con dos valores iniciales, x_l y x_u, que contienen un extremo local de $f(x)$. Después, se eligen dos puntos interiores x_1 y x_2 de acuerdo con la razón dorada,

$$d = \frac{\sqrt{5}-1}{2}(x_u - x_l)$$

$$x_1 = x_l + d$$

$$x_2 = x_u - d$$

La función se evalúa en estos dos puntos interiores. Dos casos pueden presentarse:

1. Si, como es el caso en la figura 13.4, $f(x_1) > f(x_2)$, entonces el dominio de x a la izquierda de x_2, de x_l a x_2, se puede eliminar, ya que no contiene el máximo. En este caso, x_2 será el nuevo x_l en la siguiente vuelta.
2. Si $f(x_2) > f(x_1)$, entonces el dominio de x a la derecha de x_1, de x_1 a x_u podrá eliminarse. En este caso, x_1 será el nuevo x_u en la siguiente iteración.

Ahora, ésta es la ventaja real del uso de la razón dorada. Debido a que los x_1 y x_2 originales se han escogido mediante la razón dorada, no se tienen que recalcular todos los valores de la función en la siguiente iteración. Por ejemplo, en el caso ilustrado en la figura 13.4, el anterior x_1 será el nuevo x_2. Esto significa que ya se tiene el valor para el nuevo $f(x_2)$, puesto que es el mismo valor de la función en el anterior x_1.

Para completar el algoritmo, ahora sólo se necesita determinar el nuevo x_1. Esto se realiza usando la misma proporcionalidad que antes,

$$x_1 = x_l + \frac{\sqrt{5}-1}{2}(x_u - x_l)$$

Un procedimiento similar podría usarse en el caso en que el óptimo caiga del lado izquierdo del subintervalo.

Conforme las iteraciones se repiten, el intervalo que contiene el extremo se reduce rápidamente. De hecho, en cada iteración el intervalo se reduce en un factor de la razón dorada (aproximadamente 61.8%). Esto significa que después de 10 iteraciones, el intervalo se acorta aproximadamente en 0.618^{10} o 0.008 o 0.8% de su longitud inicial. Después de 20 iteraciones, se encuentra en 0.0066%. Esta reducción no es tan buena como la que se alcanza con la bisección; aunque éste es un problema más difícil.

EJEMPLO 13.1 Búsqueda de la sección dorada

Planteamiento del problema Use la búsqueda de la sección dorada para encontrar el máximo de

$$f(x) = 2 \text{ sen } x - \frac{x^2}{10}$$

dentro del intervalo $x_l = 0$ y $x_u = 4$.

Solución Primero, se utiliza la razón dorada para crear los dos puntos interiores

$$d = \frac{\sqrt{5}-1}{2}(4-0) = 2.472$$

$$x_1 = 0 + 2.472 = 2.472$$

$$x_2 = 4 - 2.472 = 1.528$$

Se evalúa la función en los puntos interiores

$$f(x_2) = f(1.528) = 2 \text{ sen } (1.528) - \frac{1.528^2}{10} = 1.765$$

$$f(x_1) = f(2.472) = 0.63$$

Debido a que $f(x_2) > f(x_1)$, el máximo está en el intervalo definido por x_l, x_2 y x_1. Así, para el nuevo intervalo, el límite inferior sigue siendo $x_l = 0$, y x_1 será el límite superior; esto es, $x_u = 2.472$. Además, el primer valor x_2 pasa a ser el nuevo x_1; es decir, $x_1 = 1.528$. Asimismo, no se tiene que recalcular $f(x_1)$ ya que se determinó en la iteración previa como $f(1.528) = 1.765$.

Todo lo que falta es calcular el nuevo valor de d y x_2,

$$d = \frac{\sqrt{5}-1}{2}(2.472-0) = 1.528$$

$$x_2 = 2.4721 - 1.528 = 0.944$$

La evaluación de la función en x_2 es $f(0.994) = 1.531$. Como este valor es menor que el valor de la función en x_1, el máximo está en el intervalo dado por x_2, x_1 y x_u.

Si el proceso se repite, se obtienen los resultados tabulados a continuación:

i	x_l	$f(x_l)$	x_2	$f(x_2)$	x_1	$f(x_1)$	x_u	$f(x_u)$	d
1	0	0	1.5279	1.7647	2.4721	0.6300	4.0000	–3.1136	2.4721
2	0	0	0.9443	1.5310	1.5279	1.7647	2.4721	0.6300	1.5279
3	0.9443	1.5310	1.5279	1.7647	1.8885	1.5432	2.4721	0.6300	0.9443
4	0.9443	1.5310	1.3050	1.7595	1.5279	1.7647	1.8885	1.5432	0.5836
5	1.3050	1.7595	1.5279	1.7647	1.6656	1.7136	1.8885	1.5432	0.3607
6	1.3050	1.7595	1.4427	1.7755	1.5279	1.7647	1.6656	1.7136	0.2229
7	1.3050	1.7595	1.3901	1.7742	1.4427	1.7755	1.5279	1.7647	0.1378
8	1.3901	1.7742	1.4427	1.7755	1.4752	1.7732	1.5279	1.7647	0.0851

Observe que el máximo está resaltado en cada iteración. Después de ocho iteraciones, el máximo se encuentra en $x = 1.4427$ con un valor de la función 1.7755. Así, el resultado converge al valor verdadero, 1.7757, en $x = 1.4276$.

Recuerde que en la bisección (sección 5.2.1), se puede calcular un límite superior exacto para el error en cada iteración. Usando un razonamiento similar, un límite superior para la búsqueda de la sección dorada se obtiene como sigue. Una vez que se termina una iteración, el valor óptimo estará en uno de los dos intervalos. Si x_2 es el valor óptimo de la función, estará en el intervalo inferior (x_l, x_2, x_1). Si x_1 es el valor óptimo de la función, estará en el intervalo superior (x_2, x_1, x_u). Debido a que los puntos interiores son simétricos, se utiliza cualquiera de los casos para definir el error.

Observando el intervalo superior, si el valor verdadero estuviera en el extremo izquierdo, la máxima distancia al valor estimado sería

$$\begin{aligned}\Delta x_a &= x_1 - x_2\\ &= x_l + R(x_u - x_l) - x_u + R(x_u - x_l)\\ &= (x_l - x_u) + 2R(x_u - x_l)\\ &= (2R - 1)(x_u - x_l)\end{aligned}$$

o $0.236(x_u - x_l)$.

Si el valor verdadero estuviera en el extremo derecho, la máxima distancia al valor estimado sería

$$\begin{aligned}\Delta x_b &= x_u - x_1\\ &= x_u - x_l - R(x_u - x_l)\\ &= (1 - R)(x_u - x_l)\end{aligned}$$

o $0.382(x_u - x_l)$. Por lo tanto, este caso podría representar el error máximo. Este resultado después se normaliza al valor óptimo de esa iteración, $x_{\text{ópt}}$, para dar

$$\varepsilon_a = (1 - R)\left|\frac{x_u - x_l}{x_{\text{ópt}}}\right| 100\% \;.$$

Esta estimación proporciona una base para terminar las iteraciones.

En la figura 13.5*a* se presenta el pseudocódigo del algoritmo para la búsqueda de la sección dorada en la maximización. En la figura 13.5*b* se muestran las pequeñas modificaciones para convertir el algoritmo en una minimización. En ambas versiones el valor x para el óptimo se regresa como el valor de la función (*dorado*). Además, el valor de $f(x)$ óptimo se regresa como la variable $f(x)$.

Usted se preguntará por qué hemos hecho énfasis en reducir las evaluaciones de la función para la búsqueda de la sección dorada. Por supuesto, para resolver una sola optimización, la velocidad ahorrada podría ser insignificante. Sin embargo, existen dos importantes casos donde minimizar el número de evaluaciones de la función llega a ser importante. Éstos son:

1. *Muchas evaluaciones.* Hay casos donde el algoritmo de búsqueda de la sección dorada puede ser parte de muchos otros cálculos. Entonces, éste podría ser llamado muchas veces. Por lo tanto, mantener el número de evaluaciones de la función en un mínimo ofrecería dar grandes ventajas en tales casos.

```
FUNCTION Gold (xlow, xhigh, maxit, es, fx)
R = (5^0.5 - 1)/2
xℓ = xlow; xu = xhigh
iter = 1
d = R * (xu - xℓ)
x1 = xℓ + d; x2 = xu - d
f1 = f(x1)
f2 = f(x2)
IF f1 > f2 THEN                                   IF f1 < f2 THEN
  xopt = x1
  fx = f1
ELSE
  xopt = x2
  fx = f2
END IF
DO
  d = R*d
  IF f1 > f2 THEN                                 IF f1 < f2 THEN
     xℓ = x2
     x2 = x1
     x1 = xℓ+d
     f2 = f1
     f1 = f(x1)
  ELSE
     xu = x1
     x1 = x2
     x2 = xu-d
     f1 = f2
     f2 = f(x2)
  END IF
  iter = iter+1
  IF f1 > f2 THEN                                 IF f1 < f2 THEN
     xopt = x1
     fx = f1
  ELSE
     xopt = x2
     fx = f2
  END IF
  IF xopt ≠ 0. THEN
     ea = (1.-R) *ABS((xu - xℓ)/xopt) * 100.
  END IF
  IF ea ≤ es OR iter ≥ maxit EXIT
END DO
Gold = xopt
END Gold
```

a) **Maximización** *b*) **Minimización**

Figura 13.5
Algoritmo para la búsqueda de la sección dorada.

2. *Evaluaciones que toman mucho tiempo.* Por razones didácticas, se usan funciones simples en la mayoría de nuestros ejemplos. Usted deberá tener en cuenta que una función puede ser muy compleja y consumir mucho tiempo en su evaluación. Por ejemplo, en una parte posterior de este libro, se describirá cómo se utiliza la optimización para estimar los parámetros de un modelo que consiste de un sistema de ecuaciones diferenciales. En tales casos, la "función" comprende la integración del modelo que tomaría mucho tiempo. Cualquier método que minimice tales evaluaciones resultará provechoso.

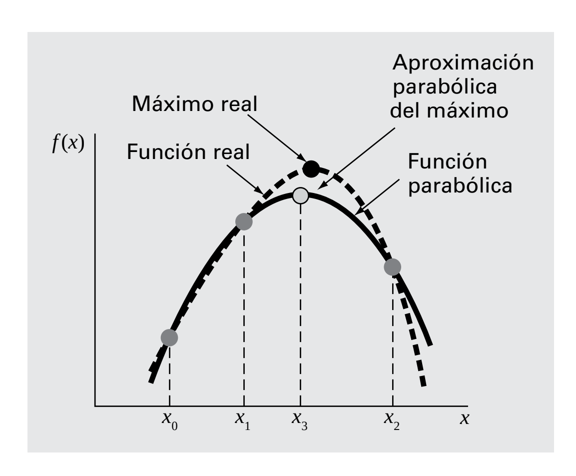


Figura 13.6
Descripción gráfica de la interpolación parabólica.

13.2 INTERPOLACIÓN PARABÓLICA

La interpolación parabólica aprovecha la ventaja de que un polinomio de segundo grado con frecuencia proporciona una buena aproximación a la forma de $f(x)$ en las cercanías de un valor óptimo (figura 13.6).

Exactamente del mismo modo en el que sólo hay una línea recta que conecte dos puntos, sólo hay una ecuación cuadrática polinómica o parábola que conecte tres puntos. De esta forma, si se tiene tres puntos que contienen un punto óptimo, se ajusta una parábola a los puntos. Después se puede derivar e igualar el resultado a cero, y así obtener una estimación de la x óptima. Es posible demostrar mediante algunas operaciones algebraicas que el resultado es

$$x_3 = \frac{f(x_0)(x_1^2 - x_2^2) + f(x_1)(x_2^2 - x_0^2) + f(x_2)(x_0^2 - x_1^2)}{2f(x_0)(x_1 - x_2) + 2f(x_1)(x_2 - x_0) + 2f(x_2)(x_0 - x_1)} \tag{13.7}$$

donde x_0, x_1 y x_2 son los valores iniciales, y x_3 es el valor de x que corresponde al valor máximo del ajuste parabólico de los valores iniciales. Después de generar el nuevo punto, hay dos estrategias para seleccionar los puntos para la siguiente iteración. El enfoque más sencillo, que es similar al método de la secante, es simplemente asignar los nuevos puntos en forma secuencial. Es decir, para la nueva iteración, $z_0 = z_1$, $z_1 = z_2$ y $z_2 = z_3$. Alternativamente, como se ilustra en el siguiente ejemplo, se puede emplear un enfoque de regionalización, similar a la bisección o la búsqueda de la sección dorada.

EJEMPLO 13.2 Interpolación parabólica

Planteamiento del problema Use la interpolación parabólica para aproximar el máximo de

$$f(x) = 2 \text{ sen } x - \frac{x^2}{10}$$

con los valores iniciales $x_0 = 0$, $x_1 = 1$ y $x_2 = 4$.

Solución Se evalúa la función en los tres valores iniciales,

$x_0 = 0$ $\quad f(x_0) = 0$
$x_1 = 1$ $\quad f(x_1) = 1.5829$
$x_2 = 4$ $\quad f(x_2) = -3.1136$

y sustituyendo en la ecuación (13.7) se obtiene,

$$x_3 = \frac{0(1^2 - 4^2) + 1.5829(4^2 - 0^2) + (-3.1136)(0^2 - 1^2)}{2(0)(1-4) + 2(1.5829)(4-0) + 2(-3.1136)(0-1)} = 1.5055$$

para la cual el valor de la función es $f(1.5055) = 1.7691$.

Después, se emplea una estrategia similar a la de la búsqueda de la sección dorada para determinar qué punto se descartará. Ya que el valor de la función en el nuevo punto es mayor que en el punto intermedio (x_1) y el nuevo valor de x está a la derecha del punto intermedio, se descarta el valor inicial inferior (x_0). Por lo tanto, para la próxima iteración,

$x_0 = 1$ $\quad f(x_0) = 1.5829$
$x_1 = 1.5055$ $\quad f(x_1) = 1.7691$
$x_2 = 4$ $\quad f(x_2) = -3.1136$

los valores se sustituyen en la ecuación (13.7) para obtener

$$x_3 = \frac{1.5829(1.5055^2 - 4^2) + 1.7691(4^2 - 1^2) + (-3.1136)(1^2 - 1.5055^2)}{2(1.5829)(1.5055 - 4) + 2(1.7691)(4-1) + 2(-3.1136)(1 - 1.5055)}$$
$$= 1.4903$$

para el cual el valor de la función es $f(1.4903) = 1.7714$.

El proceso se puede repetir, dando los resultados tabulados abajo:

i	x_0	$f(x_0)$	x_1	$f(x_1)$	x_2	$f(x_2)$	x_3	$f(x_3)$
1	0.0000	0.0000	1.0000	1.5829	4.0000	–3.1136	1.5055	1.7691
2	1.0000	1.5829	1.5055	1.7691	4.0000	–3.1136	1.4903	1.7714
3	1.0000	1.5829	1.4903	1.7714	1.5055	1.7691	1.4256	1.7757
4	1.0000	1.5829	1.4256	1.7757	1.4903	1.7714	1.4266	1.7757
5	1.4256	1.7757	1.4266	1.7757	1.4903	1.7714	1.4275	1.7757

Así, con cinco iteraciones, el resultado converge rápidamente al valor verdadero: 1.7757 en $x = 1.4276$.

Debemos mencionar que como en el método de la falsa posición, en la interpolación parabólica puede ocurrir que sólo se retenga un extremo del intervalo. Así, la convergencia puede ser lenta. Como prueba de lo anterior, observe que en nuestro ejemplo, 1.0000 fue un punto extremo en la mayoría de las iteraciones.

Este método, así como otros que usan polinomios de tercer grado, se pueden formular como parte de los algoritmos que contienen tanto pruebas de convergencia, como cuidadosas estrategias de selección para los puntos que habrán de retenerse en cada iteración y formas para minimizar la acumulación del error de redondeo.

13.3 MÉTODO DE NEWTON

Recuerde que el método de Newton-Raphson del capítulo 6 es un método abierto que permite encontrar la raíz x de una función de tal manera que $f(x) = 0$. El método se resume como

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

Se utiliza un método abierto similar para encontrar un valor óptimo de $f(x)$ al definir una nueva función, $g(x) = f'(x)$. Así, como el mismo valor óptimo x^* satisface ambas funciones

$$f'(x^*) = g(x^*) = 0$$

se emplea lo siguiente

$$x_{i+1} = x_i - \frac{f'(x_i)}{f''(x_i)} \tag{13.8}$$

como una técnica para encontrar el mínimo o máximo de $f(x)$. Se deberá observar que esta ecuación también se obtiene escribiendo una serie de Taylor de segundo grado para $f(x)$ e igualando la derivada de la serie a cero. El método de Newton es abierto y similar al de Newton-Raphson, pues no requiere de valores iniciales que contengan al óptimo. Además, también tiene la desventaja de que llega a ser divergente. Por último, usualmente es una buena idea verificar que la segunda derivada tenga el signo correcto para confirmar que la técnica converge al resultado deseado.

EJEMPLO 13.3 Método de Newton

Planteamiento del problema Con el método de Newton encuentre el máximo de

$$f(x) = 2 \operatorname{sen} x - \frac{x^2}{10}$$

con un valor inicial de $x_0 = 2.5$.

Solución La primera y segunda derivadas de la función se calculan para obtener

$$f'(x) = 2 \cos x - \frac{x}{5}$$

$$f''(x) = -2 \operatorname{sen} x - \frac{1}{5}$$

las cuales se sustituyen en la ecuación (13.8) para llegar a

Figura 13.7
Pseudocódigo para el algoritmo de Brent para hallazgo de mínimos basado en un archivo M de MATLAB desarrollado por Cleve Moler (2005).

```
Function fminsimp(x1, xu)
tol = 0.000001; phi = (1 + √5)/2; rho = 2 − phi
u = x1 + rho*(xu − x1); v = u; w = u; x = u
fu = f(u); fv = fu; fw = fu; fx = fu
xm = 0.5*(x1 + xu); d = 0; e = 0
DO
  IF |x − xm| ≤ tol EXIT
  para = |e| > tol
  IF para THEN                    (Prueba del ajuste parabólico)
    r = (x − w)*(fx − fv); q = (x − v)*(fx − fw)
    p = (x − v)*q − (x − w)*r; s = 2*(q − r)
    IF s > 0 THEN p = −p
    s = |s|
    ' Is the parabola acceptable?
    para = |p| < |0.5*s*e| And p > s*(x1 − x) And p < s*(xu − x)
    IF para THEN
      e = d; d = p/s          (Paso de interpolación parabólica)
    ENDIF
  ENDIF
  IF Not para THEN
    IF x ≥ xm THEN            (Paso de búsqueda de sección dorada)
      e = x1 − x
    ELSE
      e = xu − x
    ENDIF
    d = rho*e
  ENDIF
  u = x + d; fu = f(u)
  IF fu ≤ fx THEN             (Actualización x1, xu, x, v, w, xm)
    IF u ≥ x THEN
      x1 = x
    ELSE
      xu = x
    ENDIF
    v = w; fv = fw; w = x; fw = fx; x = u; fx = fu
  ELSE
    IF u < x THEN
      x1 = u
    ELSE
      xu = u
    ENDIF
    IF fu ≤ fw Or w = x THEN
       v = w; fv = fw; w = u; fw = fu
    ELSEIF fu ≤ fv Or v = x Or v = w THEN
       v = u; fv = fu
    ENDIF
  ENDIF
  xm = 0.5*(x1 + xu)
ENDDO
fminsimp = fu
END fminsimp
```

$$x_{i+1} = x_i - \frac{2 \cos x_i - x_i / 5}{-2 \text{ sen } x_i - 1/5}$$

Al sustituir el valor inicial se obtiene

$$x_1 = 2.5 - \frac{2 \cos 2.5 - 2.5/5}{-2 \text{ sen } 2.5 - 1/5} = 0.99508$$

para la cual el valor de la función es 1.57859. La segunda iteración da

$$x_1 = 0.995 - \frac{2 \cos 0.995 - 0.995/5}{-2 \text{ sen } 0.995 - 1/5} = 1.46901$$

que tiene como valor de la función 1.77385.

El proceso se repite, dando los resultados abajo tabulados:

i	*x*	*f*(*x*)	*f'*(*x*)	*f''*(*x*)
0	2.5	0.57194	–2.10229	–1.39694
1	0.99508	1.57859	0.88985	–1.87761
2	1.46901	1.77385	–0.09058	–2.18965
3	1.42764	1.77573	–0.00020	–2.17954
4	1.42755	1.77573	0.00000	–2.17952

Así, después de cuatro iteraciones, el resultado converge en forma rápida al valor verdadero.

Aunque el método de Newton funciona bien en algunos casos, no es práctico en otros donde las derivadas no se pueden calcular fácilmente. En tales casos, hay otros procedimientos que no implican la evaluación de la derivada. Por ejemplo, se puede desarrollar una versión del método de Newton parecida al de secante usando aproximaciones de diferencias finitas para las evaluaciones de derivada.

Una desventaja importante de este método es que llega a diverger según sea la naturaleza de la función y la calidad del valor inicial. Así, usualmente se emplea sólo cuando se está cerca del valor óptimo. Las técnicas híbridas que usan *métodos cerrados* lejos del óptimo y los *métodos abiertos* cercanos al óptimo intentan aprovechar las fortalezas de ambos procedimientos.

13.4 MÉTODO DE BRENT

Recuerde que en la sección 6.4 se describió el método de Brent para la ubicación de raíces. Este método híbrido combinaba varios métodos de búsqueda de raíces en un solo algoritmo que equilibraba confiabilidad con eficiencia.

Brent también desarrolló un enfoque similar para la minimización unidimensional. Éste combina la lenta y confiable búsqueda de la sección dorada con la interpolación parabólica, más rápida, pero posiblemente no confiable. Este procedimiento intenta primero la interpolación parabólica y sigue aplicándola mientras se obtengan resultados aceptables. De no ser así, entonces usa la búsqueda de la sección dorada para poner el asunto en orden.

La figura 13.7 presenta el pseudocódigo para el algoritmo basado en un archivo M de MATLAB desarrollado por Cleve Moler (2005). Éste representa una versión reducida a su mínima expresión de la función `fminbnd`, que es la función profesional de minimización que se utiliza en MATLAB. Por esta razón, se llama la versión simplificada `fminsimp`. Observe que necesita otra función `f` que mantiene la ecuación para la cual se está evaluando el mínimo.

Esto concluye el estudio de métodos para resolver los óptimos de las funciones de una sola variable. Algunos conceptos de ingeniería se presentan en el capítulo 16. Además, las técnicas aquí descritas son un importante elemento de algunos procedimientos para optimizar funciones con variables múltiples, como se explica en el capítulo 14.

PROBLEMAS

13.1 Dada la fórmula

$f(x) = -x^2 + 8x - 12$

a) Determine en forma analítica (esto es, por medio de derivación) el valor máximo y el correspondiente de x para esta función.
b) Verifique que la ecuación (13.7) produce los mismos resultados con base en los valores iniciales de $x_0 = 0$, $x_1 = 2$ y $x_2 = 6$.

13.2 Dada la función

$f(x) = -1.5x^6 - 2x^4 + 12x$

a) Grafique la función.
b) Utilice métodos analíticos para probar que la función es cóncava para todos los valores de x.
c) Derive la función y después use algún método de localización de raíces para resolver cuál es el máximo $f(x)$ y el valor correspondiente de x.

13.3 Encuentre el valor de x que maximiza $f(x)$ en el problema 13.2 con el uso de la búsqueda de la sección dorada. Emplee valores iniciales de $x_l = 0$ y $x_u = 2$ y realice tres iteraciones.

13.4 Repita el problema 13.3, pero use la interpolación parabólica del mismo modo que en el ejemplo 13.2. Use valores iniciales $x_0 = 0$, $x_1 = 1$ y $x_2 = 2$, y realice tres iteraciones.

13.5 Repita el problema 13.3 pero use el método de Newton. Utilice un valor inicial de $x_0 = 2$ y lleve a cabo tres iteraciones.

13.6 Emplee los métodos siguientes para encontrar el máximo de

$f(x) = 4x - 1.8x^2 + 1.2x^3 - 0.3x^4$

a) Búsqueda de la sección dorada ($x_l = -2$, $x_u = 4$, $\varepsilon_s = 1\%$).
b) Interpolación parabólica ($x_0 = 1.75$, $x_1 = 2$, $x_2 = 2.5$, iteraciones = 4). Seleccione nuevos puntos secuencialmente como en el método de la secante.
c) Método de Newton ($x_0 = 3$, $\varepsilon_s = 1\%$).

13.7 Considere la función siguiente:

$f(x) = -x^4 - 2x^3 - 8x^2 - 5x$

Use los métodos analítico y gráfico para demostrar que la función tiene un máximo para algún valor de x en el rango $-2 \le x \le 1$.

13.8 Emplee los métodos siguientes para encontrar el máximo de la función del problema 13.7:

a) Búsqueda de la sección dorada ($x_l = -2$, $x_u = 1$, $\varepsilon_s = 1\%$).
b) Interpolación parabólica ($x_0 = -2$, $x_1 = -1$, $x_2 = 1$, iteraciones = 4). Seleccione nuevos puntos secuencialmente como en el método de la secante.
c) Método de Newton ($x_0 = -1$, $\varepsilon_s = 1\%$).

13.9 Considere la función siguiente:

$$f(x) = 2x + \frac{3}{x}$$

Realice 10 iteraciones de interpolación parabólica para ubicar el mínimo. Seleccione nuevos puntos del mismo modo que en el ejemplo 13.2. Comente sobre la convergencia de sus resultados ($x_0 = 0.1$, $x_1 = 0.5$, $x_2 = 5$).

13.10 Considere la función que sigue:

$f(x) = 3 + 6x + 5x^2 + 3x^3 + 4x^4$

Localice el mínimo por medio de encontrar la raíz de la derivada de dicha función. Utilice el método de bisección con valores iniciales de $x_l = -2$ y $x_u = 1$.

13.11 Determine el mínimo de la función del problema 13.10 con los métodos siguientes:

a) Método de Newton ($x_0 = -1$, $\varepsilon_s = 1\%$).
b) Método de Newton, pero con el uso de una aproximación en diferencias finitas para las estimaciones de las derivadas:

$$f'(x) = \frac{f(x_i + \delta x_i) - f(x_i - \delta x_i)}{2\delta x_i}$$

$$f''(x) = \frac{f(x_i + \delta x_i) - 2f(x_i) - f(x_i - \delta x_i)}{(\delta x_i)^2}$$

donde δ = fracción de perturbación (= 0.01). Use un valor inicial de $x_0 = -1$ y haga iteraciones hasta que $\varepsilon_s = 1\%$.

13.12 Desarrolle un programa con el empleo de un lenguaje de programación o de macros, para implantar el algoritmo de la búsqueda de la sección dorada. Diseñe el programa expresamente para que localice un máximo. La subrutina debe tener las características siguientes:

- Iterar hasta que el error relativo esté por debajo de un criterio de detención o exceda un número máximo de iteraciones.
- Dar los valores óptimos tanto de x como de $f(x)$.
- Minimice el número de evaluaciones de la función.

Pruebe su programa con el mismo problema del ejemplo 13.1.

13.13 Desarrolle un programa como el que se describe en el problema 13.12, pero haga que ejecute una minimización o una maximización en función de la preferencia del usuario.

13.14 Desarrolle un programa usando un lenguaje de programación o macro para implementar el algoritmo de interpolación parabólica. Diseñe el programa de modo que esté expresamente creado para ubicar un máximo, y seleccione nuevos puntos como en el ejemplo 13.2. La subrutina debe tener las siguientes características:

- Estar basada en dos valores iniciales, y hacer que el programa genere el tercer valor inicial en el punto medio del intervalo.
- Comprobar si los dos valores iniciales comprenden un máximo. Si no fuera así, la subrutina no debe ejecutar el algoritmo, sino enviar un mensaje de error.
- Iterar hasta que el error relativo esté por debajo de un criterio de terminación o exceda un número máximo de iteraciones.
- Dar los valores óptimos tanto de x como de $f(x)$.
- Minimizar el número de evaluaciones de la función.

Pruebe su programa con el mismo problema del ejemplo 13.2.

13.15 Desarrolle un programa por medio de un lenguaje de programación o de macros para implantar el método de Newton. La subrutina debe tener las características siguientes:

- Iterar hasta que el error relativo esté por debajo de un criterio de terminación o supere un número máximo de iteraciones.
- Obtener los valores óptimos tanto de x como de $f(x)$.

Pruebe su programa con el mismo problema del ejemplo 13.3.

13.16 En ciertos puntos detrás de un aeroplano se hacen mediciones de la presión. Los datos tienen el mejor ajuste con la curva $y = 6 \cos x - 1.5 \operatorname{sen} x$, desde $x = 0$ hasta 6 s. Utilice cuatro iteraciones del método de la búsqueda de la sección dorada para encontrar la presión mínima. Elija $x_l = 2$ y $x_u = 4$.

13.17 La trayectoria de una pelota se calcula por medio de la ecuación

$$y = (\tan \theta_0)x - \frac{g}{2v_0^2 \cos^2 \theta_0} x^2 + y_0$$

donde y = altura (m), θ_0 = ángulo inicial (radianes), v_0 = velocidad inicial (m/s), g = constante gravitacional = 9.81 m/s^2 y y_0 = altura

inicial (m). Use el método de la búsqueda de la sección dorada para determinar la altura máxima dado que y_0 = 1 m, v_0 = 25 m/s y θ_0 = 50°. Haga iteraciones hasta que el error aproximado esté por debajo de ε_s = 1%, con el uso de valores iniciales de x_l = 0 y x_u = 60 m.

13.18 La deflexión de una trabe uniforme sujeta a una carga con distribución creciente en forma lineal, se calcula con

$$y = \frac{w_0}{120EIL}(-x^5 + 2L^2x^3 - L^4x)$$

Dado que L = 600 cm, E = 50 000 kN/cm², I = 30 000 cm⁴ y w_0 = 2.5 kN/cm, determine el punto de deflexión máximo con los métodos *a*) gráfico, *b*) de la búsqueda de la sección dorada hasta que el error aproximado esté por debajo de ε_s = 1% con valores iniciales de x_l = 0 y $x_u = L$.

13.19 Desde la superficie de la Tierra, se lanza hacia arriba un objeto con masa de 100 kg a una velocidad de 50 m/s. Si el objeto está sujeto a un arrastre lineal (c = 15 kg/s), use el método de la búsqueda de la sección dorada para determinar la altura máxima que alcanza el objeto. *Sugerencia:* Repase la sección PT4.1.2.

13.20 La distribución normal es una curva con forma de campana definida por la ecuación

$$y = e^{-x^2}$$

Utilice el método de la búsqueda de la sección dorada para determinar la ubicación del punto de deflexión de esta curva para un valor positivo de x.

13.21 Un objeto se puede proyectar hacia arriba a una velocidad específica. Si está sujeto a arrastre lineal, su altura como función del tiempo se puede calcular como

$$z = z_0 + \frac{m}{c}\left(v_0 + \frac{mg}{c}\right)\left(1 - e^{-(c/m)t}\right) - \frac{mg}{c}t$$

donde z = altura (m) sobre la superficie terrestre (definida como z = 0), z_0 = altura inicial (m), m = masa (kg), c = coeficiente lineal de arrastre (kg/s), v_0 = velocidad inicial, m/s y t = tiempo (s). Observe que, para esta formulación, la velocidad positiva se considera en dirección ascendente. Dados los siguientes valores de parámetros: g = 9.81 m/s², z_0 = 100 m, v_0 = 55 m/s, m = 80 kg y c = 15 kg/s, se puede usar la ecuación para calcular la altura del saltador. Determine el tiempo y la altura de la elevación máxima *a*) gráficamente, *b*) analíticamente, *c*) con la búsqueda de la sección dorada hasta que el error caiga debajo ε_s = 1% con valores iniciales de t_l = 0 y t_u = 10 s.

13.22 Use la búsqueda de sección dorada para determinar la longitud de la escalera más corta que llegue del suelo sobre la cerca para alcanzar la pared del edificio (figura P13.22). Pruébela para el caso en el que $h = d$ = 4 m.

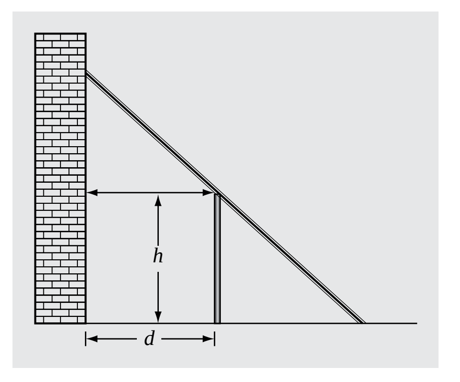


Figura P13.22
Escalera reclinada contra una reja que apenas toca la pared.

CAPÍTULO 14

Optimización multidimensional sin restricciones

Este capítulo describe las técnicas para encontrar el mínimo o el máximo de una función en varias variables. Recuerde que en el capítulo 13 nuestra imagen visual de una búsqueda unidimensional fue como una montaña rusa. En el caso de dos dimensiones, la imagen es como de montañas y valles (figura 14.1). Para problemas de más dimensiones, no es posible implementar imágenes.

Se ha optado por limitar este capítulo al caso de dos dimensiones. Esto se debe a que las características esenciales de las búsquedas multidimensionales se comunican mejor de forma visual.

Las técnicas para la optimización multidimensional sin restricciones se clasifican de varias formas. Para propósitos del presente análisis, se dividirán dependiendo de si se requiere la evaluación de la derivada. Los procedimientos que no requieren dicha evaluación se llaman *métodos sin gradiente* o *directos*. Aquellos que requieren derivadas se conocen como *métodos de gradientes* o *métodos de descenso* (o *ascenso*).

14.1 MÉTODOS DIRECTOS

Estos métodos van desde procedimientos muy burdos hasta técnicas más elegantes que intentan aprovechar la naturaleza de la función. Se empezará el análisis con un método burdo.

14.1.1 Búsqueda aleatoria

Un simple ejemplo de los métodos burdos es el *método de la búsqueda aleatoria*. Como su nombre lo indica, dicho método evalúa en forma repetida la función con los valores seleccionados aleatoriamente de la variable independiente. Si el método se lleva a cabo con un número suficiente de muestras, el óptimo eventualmente se localizará.

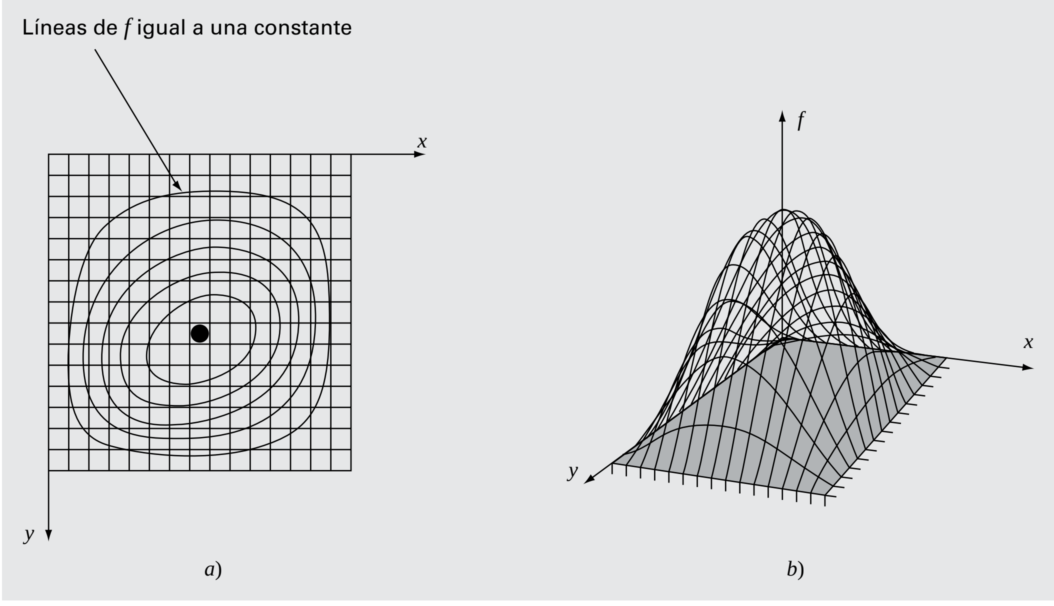


Figura 14.1
La forma más fácil de visualizar las búsquedas en dos dimensiones es en el contexto del ascenso de una montaña (maximización) o del descenso a un valle (minimización). *a*) Mapa topográfico bidimensional (2-D) de la montaña que corresponde a la gráfica tridimensional (3-D) de la montaña en el inciso *b*).

EJEMPLO 14.1 Método de la búsqueda aleatoria

Planteamiento del problema Utilice un generador de números aleatorios para localizar el máximo de

$$f(x, y) = y - x - 2x^2 - 2xy - y^2 \quad \text{(E14.1.1)}$$

en el dominio acotado por $x = -2$ a 2, y $y = 1$ a 3. El dominio se muestra en la figura 14.2. Observe que un solo máximo de 1.5 se encuentra en $x = -1$ y $y = 1.5$.

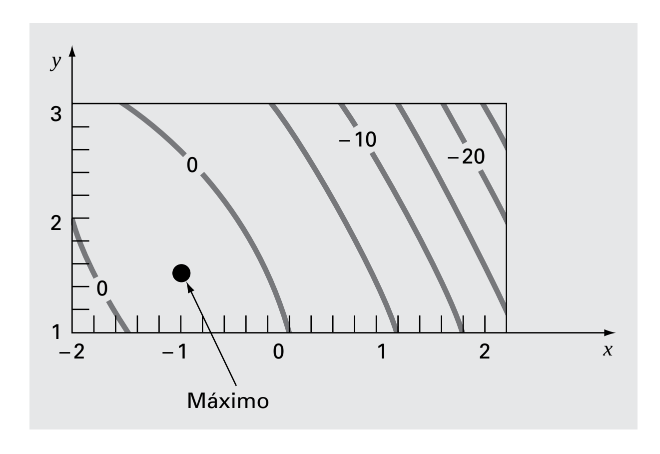


Figura 14.2
Ecuación (E14.1.1) que muestra el máximo en $x = -1$ y $y = 1.5$.

Solución Por lo común, los generadores de números aleatorios proporcionan valores entre 0 y 1. Si se designa a tal número como *r*, la siguiente fórmula se usa para generar valores de *x* aleatorios en un rango entre x_l y x_u:

$$x = x_l + (x_u - x_l)r$$

En el presente ejemplo, $x_l = -2$ y $x_u = 2$, y la fórmula es

$$x = -2 + (2 - (-2))r = -2 + 4r$$

Esto se prueba al sustituir 0 y 1 para obtener –2 y 2, respectivamente.

De manera similar para *y*, una fórmula para el mismo ejemplo se desarrolla como

$$y = y_l + (y_u - y_l)r = 1 + (3 - 1)r = 1 + 2r$$

El siguiente macrocódigo VBA de Excel utiliza la función número aleatorio Rnd de VBA, para generar un par de valores (*x*, *y*) que se sustituyen en la ecuación (E.14.1.1). El valor máximo obtenido en estos ensayos aleatorios se guarda en la variable *maxf*, y los valores correspondientes de *x* y *y* en *maxx* y *maxy*, respectivamente.

```
maxf = –1E9
For j = 1 To n
  x = –2 + 4 * Rnd
  y = 1 + 2 * Rnd
  fn = y – x – 2 * x ^ 2 – 2 * x * y – y ^ 2
  If fn > maxf Then
    maxf = fn
    maxx = x
    maxy = y
  End If
Next j
```

Después de varias iteraciones se obtiene

Iteraciones	*x*	*y*	*f*(*x*, *y*)
1000	–0.9886	1.4282	1.2462
2000	–1.0040	1.4724	1.2490
3000	–1.0040	1.4724	1.2490
4000	–1.0040	1.4724	1.2490
5000	–1.0040	1.4724	1.2490
6000	–0.9837	1.4936	1.2496
7000	–0.9960	1.5079	1.2498
8000	–0.9960	1.5079	1.2498
9000	–0.9960	1.5079	1.2498
10000	–0.9978	1.5039	1.2500

Los resultados indican que la técnica permite encontrar rápidamente el máximo verdadero.

Este simple procedimiento burdo funciona aun en discontinuidades y funciones no diferenciables. Además, siempre encuentra el óptimo global más que el local. Su principal deficiencia es que conforme crece el número de variables independientes, la implementación requerida llega a ser costosa. Además, no es eficiente, ya que no toma en cuenta el comportamiento de la función. Los procedimientos siguientes descritos en este capítulo sí toman en cuenta el comportamiento de la función, así como los resultados de las iteraciones previas para mejorar la velocidad de la convergencia. En consecuencia, aunque la búsqueda aleatoria puede probar ser útil en un contexto de problemas específico, los siguientes métodos tienen una utilidad más general y casi siempre tienen la ventaja de lograr una convergencia más eficiente.

Debemos hacer notar que se dispone de técnicas de búsqueda más sofisticadas. Éstas constituyen procedimientos heurísticos que fueron desarrollados para resolver problemas no lineales y/o discontinuos, que la optimización clásica usualmente no maneja bien. La simulación del templado de materiales, la búsqueda tabú, las redes neuronales artificiales y los algoritmos genéticos son unos pocos ejemplos. El más ampliamente utilizado es el *algoritmo genético*, en un número considerable de paquetes comerciales. En Holland (1975), iniciador del procedimiento del algoritmo genético, y Davis (1991) y Goldberg (1989) se encuentra un buen repaso de la teoría y la aplicación del método.

Figura 14.3
Descripción gráfica de cómo se presenta una búsqueda univariable.

14.1.2 Búsquedas univariables y búsquedas patrón

Es muy agradable tener un procedimiento de optimización eficiente que no requiera evaluar las derivadas. El método de búsqueda aleatoria, previamente descrito, no requiere la evaluación de la derivada, pero no es muy eficiente. En esta sección se describe un procedimiento, el método de búsqueda univariable, que es más eficiente y además no requiere la evaluación de la derivada.

La estrategia básica del *método de búsqueda univariable* consiste en trabajar sólo con una variable a la vez, para mejorar la aproximación, mientras las otras se mantienen constantes. Puesto que únicamente cambia una variable, el problema se reduce a una secuencia de búsquedas en una dimensión, que se resuelven con una diversidad de métodos (dentro de ellos, los descritos en el capítulo 13).

Realicemos una búsqueda univariable por medio de una gráfica, como se muestra en la figura 14.3. Se comienza en el punto 1, y se mueve a lo largo del eje x con y constante hacia el máximo en el punto 2. Se puede ver que el punto 2 es un máximo, al observar que la trayectoria a lo largo del eje x toca justo una línea de contorno en ese punto. Luego, muévase a lo largo del eje y con x constante hacia el punto 3. Continúa este proceso generándose los puntos 4, 5, 6, etcétera.

Aunque se está moviendo en forma gradual hacia el máximo, la búsqueda comienza a ser menos eficiente al moverse a lo largo de una cresta angosta hacia el máximo. Sin embargo, también observe que las líneas unen puntos alternados como 1-3, 3-5 o 2-4, 4-6 que van en la dirección general del máximo. Esas trayectorias presentan una oportunidad para llegar directamente a lo largo de la cresta hacia el máximo. Dichas trayectorias se denominan *direcciones patrón*.

Hay algoritmos formales que capitalizan la idea de las direcciones patrón para encontrar los valores óptimos de manera eficiente. El más conocido de tales algoritmos es el *método de Powell*, el cual se basa en la observación (véase la figura 14.4) de que si

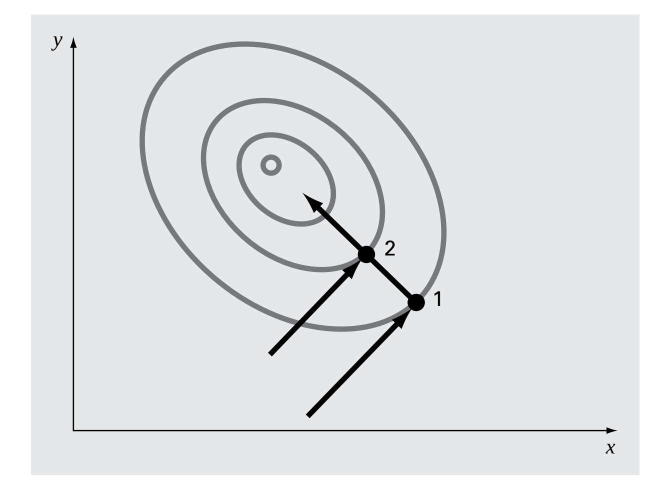


Figura 14.4
Direcciones conjugadas.

los puntos 1 y 2 se obtienen por búsquedas en una dimensión en la misma dirección, pero con diferentes puntos de partida, entonces la línea formada por 1 y 2 estará dirigida hacia el máximo. Tales líneas se llaman *direcciones conjugadas*.

En efecto, se puede demostrar que si $f(x, y)$ es una función cuadrática, las búsquedas secuenciales a lo largo de las direcciones conjugadas convergerán exactamente en un número finito de pasos, sin importar el punto de partida. Puesto que una función no lineal a menudo llega a ser razonablemente aproximada por una función cuadrática, los métodos basados en direcciones conjugadas son, por lo común, bastante eficientes y de hecho son convergentes en forma cuadrática conforme se aproximan al óptimo.

Se implementará en forma gráfica una versión simplificada del método de Powell para encontrar el máximo de

$$f(x, y) = c - (x - a)^2 - (y - b)^2$$

donde a, b y c son constantes positivas. Esta ecuación representa contornos circulares en el plano x, y, como se muestra en la figura 14.5.

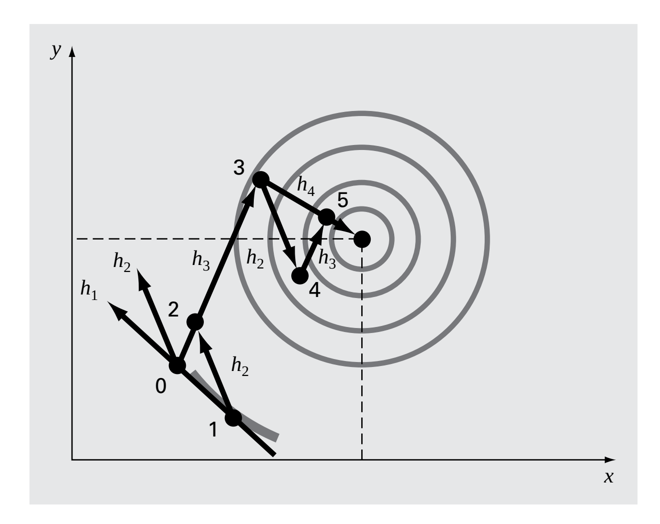


Figura 14.5
Método de Powell.

Se inicia la búsqueda en el punto cero con las direcciones iniciales h_1 y h_2. Observe que h_1 y h_2 no son necesariamente direcciones conjugadas. Desde cero, se mueve a lo largo de la dirección h_1 hasta un máximo que se localiza en el punto 1. Después se busca el punto 1 a lo largo de la dirección h_2 para encontrar el punto 2. Luego, se forma una nueva dirección de búsqueda h_3 a través de los puntos 0 y 2. Se busca a lo largo de esta dirección hasta que se localice el máximo en el punto 3. Después la búsqueda va del punto tres en la dirección h_2 hasta que se localice el máximo en el punto 4. Del punto 4 se llega al punto 5, buscando de nuevo h_3. Ahora, observe que ambos puntos, 5 y 3, se han localizado por búsqueda en la dirección h_3, desde dos puntos diferentes. Powell ha demostrado que h_4 (formado por los puntos 3 y 5) y h_3 son direcciones conjugadas. Así, buscando desde el punto 5 a lo largo de h_4, nos llevará directamente al máximo.

El método de Powell se puede refinar para volverse más eficiente; pero los algoritmos formales van más allá del alcance de este texto. Sin embargo, es un método eficiente que converge en forma cuadrática sin requerir evaluación de la derivada.

14.2 MÉTODOS CON GRADIENTE

Como su nombre lo indica, los *métodos con gradiente* utilizan en forma explícita información de la derivada para generar algoritmos eficientes que localicen el óptimo. Antes de describir los procedimientos específicos, primero se repasarán algunos conceptos y operaciones matemáticos clave.

14.2.1 Gradientes y hessianos

Recuerde del cálculo que la primera derivada de una función unidimensional proporciona la pendiente de la recta tangente a la función que se analiza. Desde el punto de vista de la optimización, ésta es una información útil. Por ejemplo, si la pendiente es positiva, nos indica que al incrementar el valor de la variable independiente nos conducirá a un valor más alto de la función que se está analizando.

Del cálculo, también recuerde que la primera derivada puede indicarnos cuándo se ha encontrado un valor óptimo, puesto que éste es el punto donde la derivada toma el valor de cero. Además, el signo de la segunda derivada puede indicarnos si se ha alcanzado un mínimo (positivo en la segunda derivada) o un máximo (negativo en la segunda derivada).

Esas ideas fueron útiles en los algoritmos de búsqueda en una dimensión que se estudiaron en el capítulo anterior. No obstante, para entender por completo las búsquedas multidimensionales, se debe primero entender cómo se expresan la primera y la segunda derivada en un contexto multidimensional.

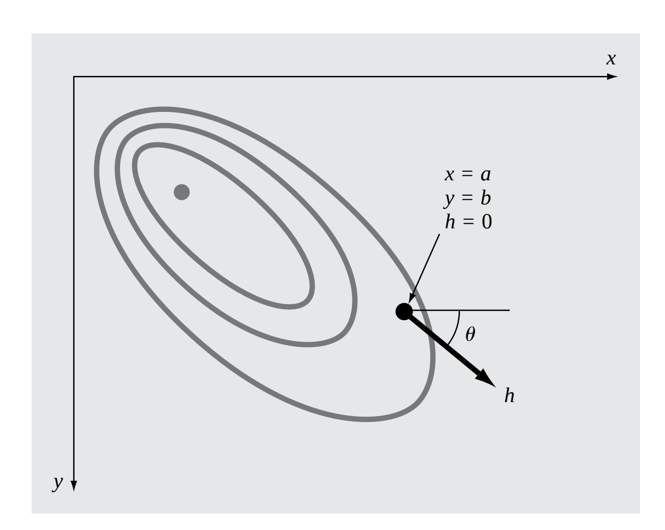


Figura 14.6
El gradiente direccional se define a lo largo de un eje *h* que forma un ángulo θ con el eje *x*.

El gradiente Suponga que se tiene una función en dos dimensiones $f(x, y)$. Un ejemplo podría ser su altura sobre una montaña como una función de su posición. Suponga que usted está en un lugar específico sobre la montaña (a, b) y quiere conocer la pendiente en una dirección arbitraria. Una forma de definir la dirección es a lo largo de un nuevo eje *h* que forma un ángulo θ con el eje *x* (figura 14.6). La elevación a lo largo de un nuevo eje puede entenderse como una nueva función $g(h)$. Si usted define su posición como el origen de este eje (es decir, $h = 0$), la pendiente en esta dirección podría designarse como $g'(0)$. Esta pendiente, que se llama *derivada direccional*, se puede calcular a partir de las derivadas parciales a lo largo de los ejes *x* y *y* mediante

$$g'(0) = \frac{\partial f}{\partial x} \cos \theta + \frac{\partial f}{\partial y} \operatorname{sen} \theta \tag{14.1}$$

donde las derivadas parciales son evaluadas en $x = a$ y $y = b$.

Suponiendo que su objetivo es obtener la mayor elevación con el siguiente paso, ahora la pregunta lógica sería: ¿en qué dirección está el mayor paso de ascenso? La respuesta a esta pregunta es proporcionada mediante lo que matemáticamente se conoce como el *gradiente*, el cual se define así:

$$\nabla f = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j} \tag{14.2}$$

Este vector también se conoce como “nabla *f*”, el cual se relaciona con la derivada direccional de $f(x, y)$ en el punto $x = a$ y $y = b$.

La notación vectorial ofrece un medio conciso para generalizar el gradiente a *n* dimensiones,

$$\nabla f(\mathbf{x}) = \begin{Bmatrix} \dfrac{\partial f}{\partial x_1}(\mathbf{x}) \\ \dfrac{\partial f}{\partial x_2}(\mathbf{x}) \\ \cdot \\ \cdot \\ \cdot \\ \dfrac{\partial f}{\partial x_n}(\mathbf{x}) \end{Bmatrix}$$

¿Cómo se usa el gradiente? Para el problema de subir la montaña, si lo que interesa es ganar elevación tan rápidamente como sea posible, el gradiente nos indica, de manera local, qué dirección tomar y cuánto ganaremos al hacerlo. Observe, sin embargo, que dicha estrategia ¡no necesariamente nos lleva en una trayectoria directa a la cima! Más tarde, en este capítulo, se analizarán estas ideas con mayor profundidad.

EJEMPLO 14.2 Utilización del gradiente para evaluar la trayectoria de máxima pendiente

Planteamiento del problema Con el gradiente evalúe la dirección de máxima pendiente para la función

$$f(x, y) = xy^2$$

en el punto (2, 2). Se considera que la *x* positiva está dirigida hacia el este y la *y* positiva hacia el norte.

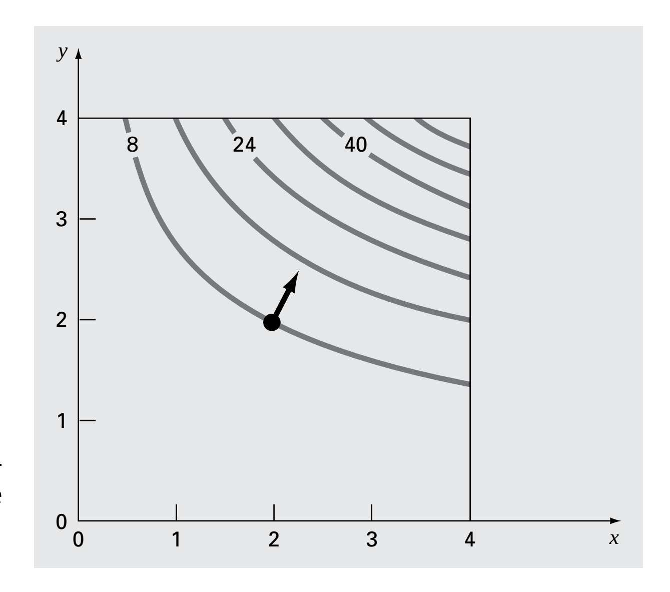


Figura 14.7
La flecha sigue la dirección del ascenso de mayor pendiente calculado con el gradiente.

Solución Primero, la elevación se determina así

$$f(2, 2) = 2(2)^2 = 8$$

Ahora, se evalúan las derivadas parciales,

$$\frac{\partial f}{\partial x} = y^2 = 2^2 = 4$$

$$\frac{\partial f}{\partial y} = 2xy = 2(2)(2) = 8$$

las cuales se usan para determinar el gradiente como

$$\nabla f = 4\mathbf{i} + 8\mathbf{j}$$

Este vector puede bosquejarse en un mapa topográfico de la función, como en la figura 14.7. Esto inmediatamente nos indica que la dirección que debe tomarse es

$$\theta = \tan^{-1}\left(\frac{8}{4}\right) = 1.107 \text{ radianes } (= 63.4°)$$

respecto al eje x. La pendiente en esta dirección, que es la magnitud de ∇f, se calcula así

$$\sqrt{4^2 + 8^2} = 8.944$$

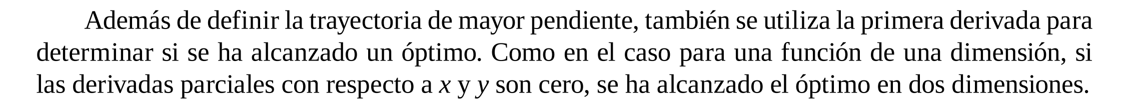
Así, durante el primer paso, inicialmente se ha ganado 8.944 unidades de aumento de elevación por unidad de distancia recorrida a lo largo de esta trayectoria con la mayor pendiente. Observe que la ecuación (14.1) da el mismo resultado,

$$g'(0) = 4\cos(1.107) + 8\,\text{sen}(1.107) = 8.944$$

Observe que para cualquier otra dirección, digamos $\theta = 1.107/2 = 0.5235$, $g'(0) = 4\cos(0.5235) + 8\,\text{sen}(0.5235) = 7.608$, que es menor.

Conforme se mueve hacia adelante, cambiarán tanto la dirección como la magnitud de la trayectoria de mayor pendiente. Estos cambios se pueden cuantificar a cada paso mediante el gradiente y la dirección del ascenso se modificará de acuerdo con ello.

Se puede obtener una mejor comprensión al inspeccionar la figura 14.7. Como se indica, la dirección de ascenso con mayor pendiente es perpendicular, u *ortogonal*, al contorno en la elevación en la coordenada (2, 2). Ésta es una propiedad del gradiente.



Además de definir la trayectoria de mayor pendiente, también se utiliza la primera derivada para determinar si se ha alcanzado un óptimo. Como en el caso para una función de una dimensión, si las derivadas parciales con respecto a x y y son cero, se ha alcanzado el óptimo en dos dimensiones.

El hessiano En problemas de una dimensión, tanto la primera como la segunda derivada ofrecen información valiosa en la búsqueda del óptimo. La primera derivada *a*) proporciona una trayectoria de máxima inclinación de la función y *b*) indica que se ha alcanzado el óptimo. Una vez en el óptimo, la segunda derivada indicará si es un máximo [$f''(x)$ negativo] o un mínimo [$f''(x)$ positivo]. En los párrafos anteriores, se ilustró cómo el gradiente proporciona la mejor trayectoria en problemas multidimensionales. Ahora, se examinará cómo se usa la segunda derivada en este contexto.

Puede esperarse que si las segundas derivadas parciales respecto de x y y son negativas ambas, entonces se ha alcanzado un máximo. La figura 14.8 muestra una función en la que esto no es cierto. El punto (a, b) de esta gráfica parece ser un mínimo cuando se observa a lo largo ya sea de la dimensión x o de la y. En ambos casos, las segundas derivadas parciales son positivas. Sin embargo, si la función se observa a lo largo de la línea $y = x$, puede verse que se presenta un máximo en el

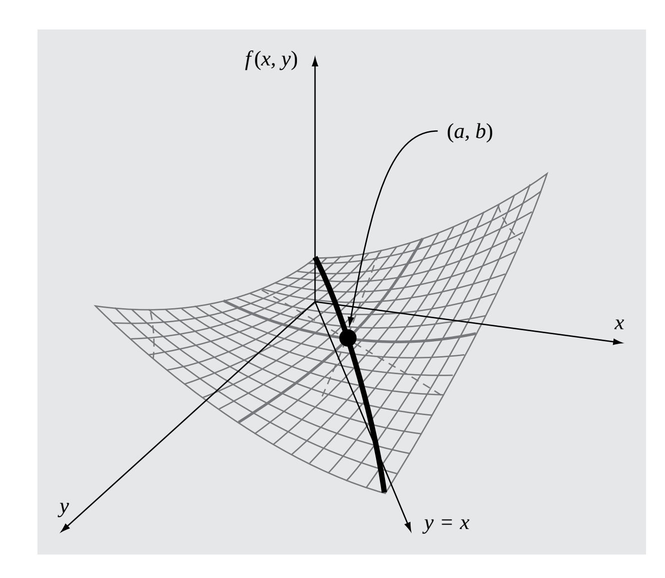


Figura 14.8
Un punto silla ($x = a$ y $y = b$). Observe que al ser vista la curva a lo largo de las direcciones x y y, parece que la función pasa por un mínimo (la segunda derivada es positiva); mientras que al verse a lo largo del eje $x = y$, es cóncava hacia abajo (la segunda derivada es negativa).

mismo punto. Éste se conoce como punto *silla* y, claramente, no se presentan ni un máximo ni un mínimo en ese punto.

Ya sea que ocurra un máximo o un mínimo, esto involucra no sólo a las primeras derivadas parciales con respecto a x y y, sino también a la segunda derivada parcial respecto a x y y. Suponiendo que las derivadas parciales sean continuas en y cerca del punto que se habrá de evaluar, se puede calcular la siguiente cantidad:

$$|H| = \frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 \tag{14.3}$$

Pueden presentarse tres casos:

- Si $|H| > 0$ y $\partial^2 f/\partial x^2 > 0$, entonces $f(x, y)$ tiene un mínimo local.
- Si $|H| > 0$ y $\partial^2 f/\partial x^2 < 0$, entonces $f(x, y)$ tiene un máximo local.
- Si $|H| < 0$, ento.nces $f(x, y)$ tiene un punto silla.

La cantidad $|H|$ es igual al determinante de una matriz formada con las segundas derivadas,[1]

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} \tag{14.4}$$

donde a esta matriz se le conoce formalmente como el *hessiano* de f.

Además de proporcionar un medio para discriminar si una función multidimensional ha alcanzado el óptimo, el hessiano tiene otros usos en optimización (por ejemplo, en la forma multidimensional del método de Newton). En particular, permite búsquedas que incluyen curvatura de segundo orden para obtener mejores resultados.

Aproximaciones por diferencias finitas Se debe mencionar que en los casos donde es difícil o inconveniente calcular analíticamente tanto el gradiente como el determinante hessiano, éstos se pueden evaluar numéricamente. En la mayoría de los casos se emplea el método que se presentó en la sección 6.3.3 para el método de la secante modificado. Es decir, las variables independientes se modifican ligeramente para generar las derivadas parciales requeridas. Por ejemplo, si se adopta el procedimiento de diferencias centrales, éstas se calculan así

$$\frac{\partial f}{\partial x} = \frac{f(x+\delta x, y) - f(x-\delta x, y)}{2\delta x} \tag{14.5}$$

$$\frac{\partial f}{\partial y} = \frac{f(x, y+\delta y) - f(x, y-\delta y)}{2\delta y} \tag{14.6}$$

$$\frac{\partial^2 f}{\partial x^2} = \frac{f(x+\delta x, y) - 2f(x, y) + f(x-\delta x, y)}{\delta x^2} \tag{14.7}$$

[1] Observe que $\partial^2 f/(\partial x \partial y) = \partial^2 f/(\partial y \partial x)$.

$$\frac{\partial^2 f}{\partial y^2} = \frac{f(x, y+\delta y) - 2f(x, y) + f(x, y-\delta y)}{\delta y^2} \tag{14.8}$$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{f(x+\delta x, y+\delta y) - f(x+\delta x, y-\delta y) - f(x-\delta x, y+\delta y) + f(x-\delta x, y-\delta y)}{4\delta x \delta y} \tag{14.9}$$

donde δ es un valor fraccional muy pequeño.

Observe que los métodos empleados en paquetes de software comerciales también usan diferencias hacia adelante. Además, son usualmente más complicados que las aproximaciones enlistadas en las ecuaciones (14.5) a la (14.9). Dennis y Schnabel (1996) dan más detalles sobre este método.

Sin importar cómo se implemente la aproximación, la cuestión importante es que se pueda tener la opción de evaluar el gradiente y/o el hessiano en forma analítica. Esto algunas veces puede resultar una tarea ardua; pero el comportamiento del algoritmo puede ser benéfico para que el esfuerzo valga la pena. Las derivadas de forma cerrada serán exactas; pero lo más importante es que se reduce el número de evaluaciones de la función. Este último detalle tiene un impacto significativo en el tiempo de ejecución.

Por otro lado, usted practicará con frecuencia la opción de calcular estas cantidades internamente mediante procedimientos numéricos. En muchos casos, el comportamiento será el adecuado y se evitará el problema de numerosas derivaciones parciales. Tal podría ser el caso de los optimizadores utilizados en ciertas hojas de cálculo y paquetes de software matemático (por ejemplo, Excel). En dichos casos, quizá no se le dé la opción de introducir un gradiente y un hessiano derivados en forma analítica. Sin embargo, para problemas de tamaño pequeño o moderado esto no representa un gran inconveniente.

14.2.2 Método de máxima inclinación

Una estrategia obvia para subir una colina sería determinar la pendiente máxima en la posición inicial y después comenzar a caminar en esa dirección. Pero claramente surge otro problema casi de inmediato. A menos que usted realmente tenga suerte y empiece sobre una cuesta que apunte directamente a la cima, tan pronto como se mueva su camino diverge en la dirección de ascenso con máxima inclinación.

Al darse cuenta de este hecho, usted podría adoptar la siguiente estrategia. Avance una distancia corta a lo largo de la dirección del gradiente. Luego deténgase, reevalúe el gradiente y camine otra distancia corta. Mediante la repetición de este proceso podrá llegar a la punta de la colina.

Aunque tal estrategia parece ser superficialmente buena, no es muy práctica. En particular, la evaluación continua del gradiente demanda mucho tiempo en términos de cálculo. Se prefiere un método que consista en moverse por un camino fijo, a lo largo del gradiente inicial hasta que $f(x, y)$ deje de aumentar; es decir, tienda a nivelarse en su dirección de viaje. Este punto se convierte en el punto inicial donde se reevalúa ∇f y se sigue una nueva dirección. El proceso se repite hasta que se alcance la cima. Este procedimiento se conoce como *método de máxima inclinación*.[2] Es la más directa de las técnicas de búsqueda con gradiente. La idea básica detrás del procedimiento se describe en la figura 14.9.

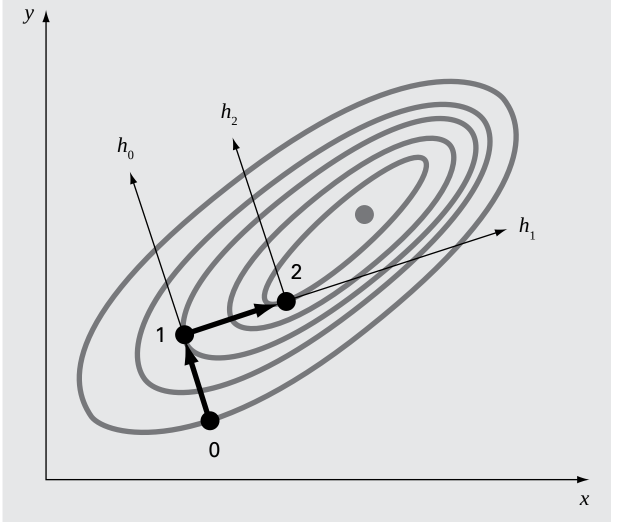


Figura 14.9
Descripción gráfica del método de máxima inclinación.

[2] Debido a nuestro énfasis sobre maximización, aquí se utiliza la terminología de *ascenso de máxima inclinación*. El mismo enfoque se puede utilizar también para la minimización; en este caso se usará la terminología de *descenso de máxima inclinación*.

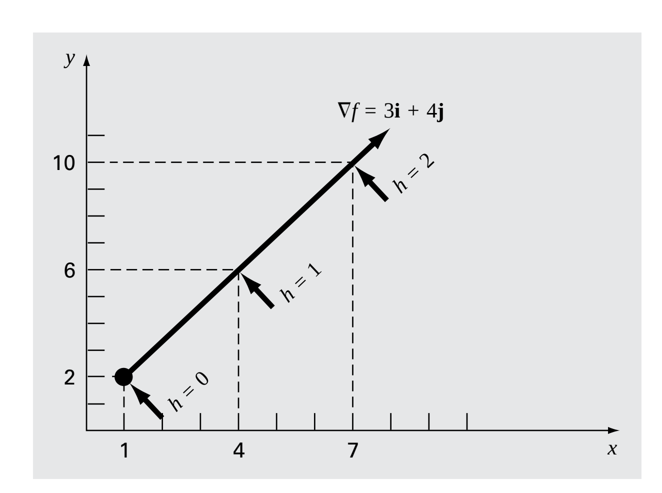


Figura 14.10
Relación entre una dirección arbitraria *h* y las coordenadas *x* y *y*.

Comenzaremos en un punto inicial (x_0, y_0) etiquetado como "0" en la figura. En este punto, se determina la dirección de ascenso con máxima inclinación; es decir, el gradiente. Entonces se busca a lo largo de la dirección del gradiente, h_0, hasta que se encuentra un máximo, que se marca como "1" en la figura. Después el proceso se repite.

Así, el problema se divide en dos partes: 1) se determina la "mejor" dirección para la búsqueda y 2) se determina "el mejor valor" a lo largo de esa dirección de búsqueda. Como se verá, la efectividad de los diversos algoritmos descritos en las siguientes páginas depende de qué tan hábiles seamos en ambas partes.

Por ahora, el método del ascenso con máxima inclinación usa el gradiente como su elección para la "mejor" dirección. Se ha mostrado ya cómo se evalúa el gradiente en el ejemplo 14.1. Ahora, antes de examinar cómo se construye el algoritmo para localizar el máximo a lo largo de la dirección de máxima inclinación, se debe hacer una pausa para explorar el modo de transformar una función de *x* y *y* en una función de *h* a lo largo de la dirección del gradiente.

Comenzando en x_0 y y_0 las coordenadas de cualquier punto en la dirección del gradiente se expresan como

$$x = x_0 + \frac{\partial f}{\partial x} h \tag{14.10}$$

$$y = y_0 + \frac{\partial f}{\partial y} h \tag{14.11}$$

donde *h* es la distancia a lo largo del eje *h*. Por ejemplo, suponga que $x_0 = 1$, $y_0 = 2$ y $\nabla f = 3\mathbf{i} + 4\mathbf{j}$, como se muestra en la figura 14.10. Las coordenadas de cualquier punto a lo largo del eje *h* están dadas por

$$x = 1 + 3h \tag{14.12}$$

$$y = 2 + 4h \tag{14.13}$$

El siguiente ejemplo ilustra la forma en que se emplean tales transformaciones para convertir una función bidimensional de *x* y *y* en una función unidimensional de *h*.

EJEMPLO 14.3 Desarrollo de una función 1-D a lo largo de la dirección del gradiente

Planteamiento del problema Suponga que se tiene la siguiente función en dos dimensiones:

$$f(x, y) = 2xy + 2x - x^2 - 2y^2$$

Desarrolle una versión unidimensional de esta ecuación a lo largo de la dirección del gradiente en el punto donde $x = -1$ y $y = 1$.

Solución Las derivadas parciales se evalúan en (–1, 1),

$$\frac{\partial f}{\partial x} = 2y + 2 - 2x = 2(1) + 2 - 2(-1) = 6$$

$$\frac{\partial f}{\partial y} = 2x - 4y = 2(-1) - 4(1) = -6$$

Por lo tanto, el vector gradiente es

$$\nabla f = 6\mathbf{i} - 6\mathbf{j}$$

Para encontrar el máximo, se busca en la dirección del gradiente; es decir, a lo largo de un eje *h* que corre en la dirección de este vector. La función se expresa a lo largo de este eje como

$$f\left(x_0 + \frac{\partial f}{\partial x}h, y_0 + \frac{\partial f}{\partial y}h\right) = f(-1+6h, 1-6h)$$

$$= 2(-1+6h)(1-6h) + 2(-1+6h) - (-1+6h)^2 - 2(1-6h)^2$$

donde las derivadas parciales se evalúan en $x = -1$ y $y = 1$.

Al combinar términos, se obtiene una función unidimensional $g(h)$ que transforma $f(x, y)$ a lo largo del eje *h*,

$$g(h) = -180h^2 + 72h - 7$$

Ahora que se ha obtenido una función a lo largo de la trayectoria de ascenso de máxima inclinación, es posible explorar cómo contestar la segunda pregunta. Esto es, ¿qué tan lejos se llega a lo largo de este camino? Un procedimiento sería moverse a lo largo de este camino hasta encontrar el máximo de la función. Identificaremos la localización de este máximo como h^*. Éste es el valor del paso que maximiza *g* (y, por lo tanto, *f*) en la dirección del gradiente. Este problema es equivalente a encontrar el máximo de una función de una sola variable *h*. Lo cual se realiza mediante diferentes técnicas de búsqueda unidimensional como las analizadas en el capítulo 13. Así, se pasa de encontrar el óptimo de una función de dos dimensiones a realizar una búsqueda unidimensional a lo largo de la dirección del gradiente.

Este método se llama *ascenso de máxima inclinación* cuando se utiliza un tamaño de paso arbitrario *h*. Si se encuentra que un valor de un solo paso h^* nos lleva directamente al máximo a lo largo de la dirección del gradiente, el método se llama *ascenso óptimo de máxima inclinación*.

EJEMPLO 14.4 Ascenso óptimo de máxima inclinación

Planteamiento del problema Maximice la siguiente función:

$$f(x, y) = 2xy + 2x - x^2 - 2y^2$$

usando los valores iniciales, $x = -1$ y $y = 1$.

Solución Debido a que esta función es muy simple, se obtiene primero una solución analítica. Para hacerlo, se evalúan las derivadas parciales

$$\frac{\partial f}{\partial x} = 2y + 2 - 2x = 0$$

$$\frac{\partial f}{\partial y} = 2x - 4y = 0$$

De este par de ecuaciones se puede encontrar el valor óptimo, en $x = 2$ y $y = 1$. Las segundas derivadas parciales también se determinan y evalúan en el óptimo,

$$\frac{\partial^2 f}{\partial x^2} = -2$$

$$\frac{\partial^2 f}{\partial y^2} = -4$$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = 2$$

y el determinante de la matriz hessiana se calcula [ecuación (14.3)],

$$|H| = -2(-4) - 2^2 = 4$$

Por lo tanto, debido a que $|H| > 0$ y $\partial^2 f/\partial x^2 < 0$, el valor de la función $f(2, 1)$ es un máximo.

Ahora se usará el método del ascenso de máxima inclinación. Recuerde que al final del ejemplo 14.3 ya se habían realizado los pasos iniciales del problema al generar

$$g(h) = -180h^2 + 72h - 7$$

Ahora, ya que ésta es una simple parábola, se puede localizar, de manera directa, el máximo (es decir, $h = h^*$) resolviendo el problema,

$$g'(h^*) = 0$$
$$-360h^* + 72 = 0$$
$$h^* = 0.2$$

Esto significa que si se viaja a lo largo del eje h, $g(h)$ alcanza un valor mínimo cuando $h = h^* = 0.2$. Este resultado se sustituye en las ecuaciones (14.10) y (14.11) para obtener las coordenadas (x, y) correspondientes a este punto,

$$x = -1 + 6(0.2) = 0.2$$
$$y = 1 - 6(0.2) = -0.2$$

Este paso se describe en la figura 14.11 conforme el movimiento va del punto 0 al 1.

El segundo paso se implementa tan sólo al repetir el procedimiento. Primero, las derivadas parciales se evalúan en el nuevo punto inicial (0.2, –0.2) para obtener

$$\frac{\partial f}{\partial x} = 2(-0.2) + 2 - 2(0.2) = 1.2$$

$$\frac{\partial f}{\partial y} = 2(0.2) - 4(-0.2) = 1.2$$

Por lo tanto, el vector gradiente es

$$\nabla f = 1.2\mathbf{i} + 1.2\mathbf{j}$$

Esto significa que la dirección de máxima inclinación está ahora dirigida hacia arriba y hacia la derecha en un ángulo de 45° con el eje x (véase la figura 14.11). Las coordenadas a lo largo de este nuevo eje h se expresan ahora como

$$x = 0.2 + 1.2h$$
$$y = -0.2 + 1.2h$$

Al sustituir estos valores en la función se obtiene

$$f(0.2 + 1.2h, -0.2 + 1.2h) = g(h) = -1.44h^2 + 2.88h + 0.2$$

El paso h^* que nos lleva al máximo a lo largo de la dirección marcada ahora se calcula directamente como

$$g'(h^*) = -2.88h^* + 2.88 = 0$$
$$h^* = 1$$

Este resultado se sustituye en las ecuaciones (14.10) y (14.11) para obtener las coordenadas (x, y) correspondientes a este nuevo punto,

$$x = 0.2 + 1.2(1) = 1.4$$
$$y = -0.2 + 1.2(1) = 1$$

Como se describe en la figura 14.11, nos movemos a las nuevas coordenadas, marcadas como punto 2 en la gráfica, y al hacer esto nos acercamos al máximo. El procedimiento se puede repetir y se obtiene un resultado final que converge a la solución analítica, $x = 2$ y $y = 1$.

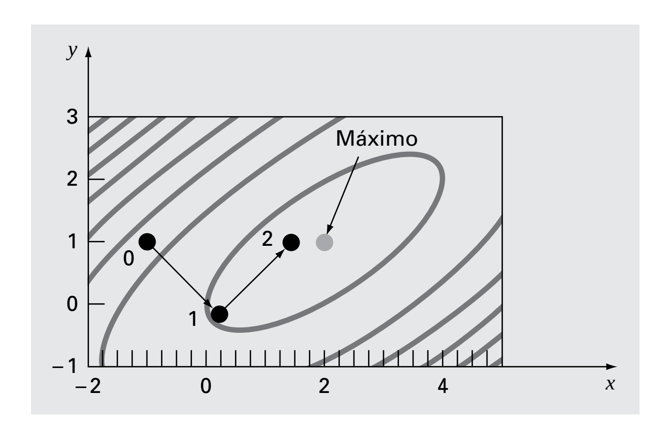


Figura 14.11
El método del ascenso óptimo de máxima inclinación.

Es posible demostrar que el método del descenso de máxima inclinación es linealmente convergente. Además, tiende a moverse de manera muy lenta, a lo largo de crestas largas y angostas. Esto porque el nuevo gradiente en cada punto máximo será perpendicular a la dirección original. Así, la técnica da muchos pasos pequeños cruzando la ruta directa hacia la cima. Por lo tanto, aunque es confiable, existen otros métodos que convergen mucho más rápido, particularmente en la vecindad de un valor óptimo. En el resto de la sección se examinan esos métodos.

14.2.3 Métodos avanzados del gradiente

Método de gradientes conjugados (Fletcher-Reeves) En la sección 14.1.2 se ha visto cómo en el método de Powell las direcciones conjugadas mejoran mucho la eficiencia de la búsqueda univariable. De manera similar, se puede también mejorar el ascenso de máxima inclinación linealmente convergente usando gradientes conjugados. En efecto, se puede demostrar que un método de optimización, que usa gradientes conjugados para definir la dirección de búsqueda, es cuadráticamente convergente. Esto también asegura que el método optimizará una función cuadrática exactamente en un número finito de pasos sin importar el punto de inicio. Puesto que la mayoría de las funciones que tienen buen comportamiento llegan a aproximarse en forma razonable bien mediante una función cuadrática en la vecindad de un óptimo, los métodos de convergencia cuadrática a menudo resultan muy eficientes cerca de un valor óptimo.

Se ha visto cómo, empezando con dos direcciones de búsqueda arbitrarias, el método de Powell produce nuevas direcciones de búsqueda conjugadas. Este método es cuadráticamente convergente y no requiere la información del gradiente. Por otro lado, si la evaluación de las derivadas es práctica, se pueden buscar algoritmos que combinen las ideas del descenso de máxima inclinación con las direcciones conjugadas, para lograr un comportamiento inicial más sólido y de convergencia rápida conforme la técnica conduzca hacia el óptimo. El *algoritmo del gradiente conjugado de Fletcher-Reeves* modifica el método de ascenso de máxima inclinación al imponer la condición de que sucesivas direcciones de búsqueda del gradiente sean mutuamente conjugadas. La prueba y el algoritmo están más allá del alcance del texto, pero se describen en Rao (1996).

Método de Newton El método de Newton para una sola variable (recuerde la sección 13.3) se puede extender a los casos multivariables. Escriba una serie de Taylor de segundo orden para $f(\mathbf{x})$ cerca de $\mathbf{x} = \mathbf{x}_i$,

$$f(\mathbf{x}) = f(\mathbf{x}_i) + \nabla f^T(\mathbf{x}_i)(\mathbf{x} - \mathbf{x}_i) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_i)^T H_i(\mathbf{x} - \mathbf{x}_i)$$

donde H_i es la matriz hessiana. En el mínimo,

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}_j} = 0 \qquad \text{para } j = 1, 2, \ldots, n$$

Así,

$$\nabla f = \nabla f(\mathbf{x}_i) + H_i(\mathbf{x} - \mathbf{x}_i) = 0$$

Si H es no singular,

$$\mathbf{x}_{i+1} = \mathbf{x}_i - H_i^{-1}\nabla f \tag{14.14}$$

la cual, se puede demostrar, converge en forma cuadrática cerca del óptimo. Este método, de nuevo, se comporta mejor que el método del ascenso de máxima inclinación (véase la figura 14.12). Sin embargo, observe que este método requiere tanto del cálculo de las segundas derivadas como de la inversión matricial, en cada iteración. Por lo que el método no es muy útil en la práctica para funciones con gran número de variables. Además, el método de Newton quizá no converja si el punto inicial no está cerca del óptimo.

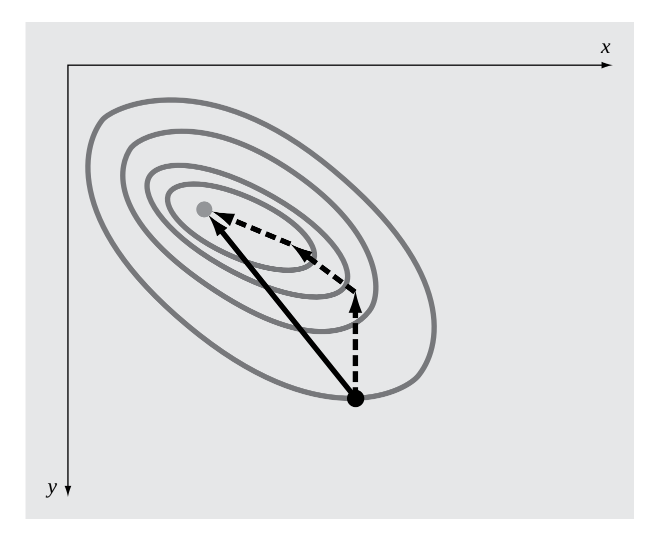


Figura 14.12
Cuando el punto inicial está cerca del punto óptimo, seguir el gradiente puede resultar ineficiente. Los métodos de Newton intentan la búsqueda a lo largo de una trayectoria directa hacia el óptimo (línea continua).

Método de Marquardt Se sabe que el método del ascenso de máxima inclinación aumenta el valor de la función, aun si el punto inicial está lejos del óptimo. Por otro lado, ya se describió el método de Newton, que converge con rapidez cerca del máximo. El *método de Marquardt* usa el método del descenso de máxima inclinación cuando **x** está lejos de **x***, y el método de Newton cuando **x** está cerca de un óptimo. Esto se puede lograr al modificar la diagonal del hessiano en la ecuación (14.14),

$$\tilde{H}_i = H_i + \alpha_i I$$

donde α_i es una constante positiva e I es la matriz identidad. Al inicio del procedimiento, se supone que α_i es grande y

$$\tilde{H}_i^{-1} \approx \frac{1}{\alpha_1} I$$

la cual reduce la ecuación (14.14) al método del ascenso de máxima inclinación. Conforme continúan las iteraciones, α_i se aproxima a cero y el método se convierte en el de Newton.

Así, el método de Marquardt ofrece lo mejor de los procedimientos: comienza en forma confiable a partir de valores iniciales pobres y luego acelera en forma rápida cuando se aproxima al óptimo. Por desgracia, el método también requiere la evaluación del hessiano y la inversión matricial en cada paso. Debe observarse que el método de Marquardt es, ante todo, útil para problemas no lineales de mínimos cuadrados.

Métodos de cuasi-Newton Los *métodos cuasi-Newton*, o *métodos de métrica variable*, buscan estimar el camino directo hacia el óptimo en forma similar al método de Newton. Sin embargo, observe que la matriz hessiana en la ecuación (14.14) se compone de las segundas derivadas de f que varían en cada paso. Los métodos cuasi-Newton intentan evitar estas dificultades al aproximar H con otra matriz A, sólo las primeras derivadas parciales de f. El procedimiento consiste en comenzar con una aproximación inicial de H^{-1} y actualizarla y mejorarla en cada iteración. Estos métodos se llaman de cuasi-Newton porque no usan el hessiano verdadero, sino más bien una aproximación. Así, se tienen dos aproximaciones simultáneas: 1) la aproximación original de la serie de Taylor y 2) la aproximación del hessiano.

Hay dos métodos principales de este tipo: los algoritmos de *Davidon-Fletcher-Powell* (DFP) y de *Broyden-Fletcher-Goldfarb-Shanno* (BFGS). Éstos son similares excepto en detalles concernientes a cómo manejan los errores de redondeo y su convergencia. BFGS es, por lo general, reconocido como superior en la mayoría de los casos. Rao (1996) proporciona detalles y declaraciones formales sobre ambos algoritmos, el DFP y el BFGS.

PROBLEMAS

14.1 Encuentre la derivada direccional de

$$f(x, y) = x^2 + 2y^2$$

si $x = 2$ y $y = 2$, en la dirección de $h = 2i + 3j$.

14.2 Repita el ejemplo 14.2 para la función siguiente, en el punto (0.8, 1.2).

$$f(x, y) = 2xy + 1.5y - 1.25x^2 - 2y^2 + 5$$

14.3 Dada

$$f(x, y) = 2.25xy + 1.75y - 1.5x^2 - 2y^2$$

construya y resuelva un sistema de ecuaciones algebraicas lineales que maximice $f(x)$. Observe que esto se logra por medio de igualar a cero las derivadas parciales de f con respecto tanto de x como de y.

14.4 *a*) Comience con un valor inicial de $x = 1$ y $y = 1$, y realice dos aplicaciones del método de ascenso de máxima inclinación para $f(x, y)$, como en el problema 14.3.

b) Haga una gráfica de los resultados del inciso *a*), en la que se muestre la trayectoria de la búsqueda.

14.5 Encuentre el vector gradiente y la matriz hessiana para cada una de las funciones siguientes:

a) $f(x, y) = 2xy^2 + 3e^{xy}$

b) $f(x, y, z) = x^2 + y^2 + 2z^2$

c) $f(x, y) = \ln(x^2 + 2xy + 3y^2)$

14.6 Encuentre el valor mínimo de

$$f(x, y) = (x - 3)^2 + (y - 2)^2$$

comience con $x = 1$ y $y = 1$, utilice el método de descenso de máxima inclinación con un criterio de detención de $\varepsilon_s = 1\%$. Explique sus resultados.

14.7 Efectúe una iteración del método de ascenso de máxima inclinación para localizar el máximo de

$$f(x, y) = 6.4x + 2y + x^2 - 2x^4 + 2xy - 3y^2$$

con los valores iniciales de $x = 0$ y $y = 0$. Emplee bisección para encontrar el tamaño óptimo de paso en la dirección de búsqueda del gradiente.

14.8 Realice una iteración del método de descenso de máxima inclinación del gradiente óptimo, para localizar el mínimo de

$$f(x, y) = -8x + x^2 + 12y + 4y^2 - 2xy$$

utilice valores iniciales de $x = 0$ y $y = 0$.

14.9 Con un lenguaje de programación o de macros, desarrolle un programa para implantar el método de búsqueda aleatoria. Diseñe el subprograma de modo que en forma expresa localice un máximo. Pruebe el programa con $f(x, y)$ del problema 14.7. Utilice un rango de –2 a 2 tanto para x como para y.

14.10 La búsqueda por malla es otro procedimiento burdo para optimizar. En la figura P14.10 se ilustra la versión para dos dimensiones. Las dimensiones x y y se dividen en incrementos a fin de formar una malla. Después, se evalúa la función en cada nodo de la malla. Entre más densa sea la malla más probable será la localización del óptimo.

Con un lenguaje de programación o de macros, desarrolle un programa para implantar el método de búsqueda por malla. Diseñe el programa expresamente para que localice un máximo. Pruébelo con el mismo problema del ejemplo 14.1.

14.11 Desarrolle una ecuación unidimensional en la dirección del gradiente de presión en el punto (4, 2). La función de presión es

$$f(x, y) = 6x^2y - 9y^2 - 8x^2$$

14.12 Una función de temperatura es

$$f(x, y) = 2x^3y^2 - 7xy - x^2 + 3y$$

Desarrolle una función unidimensional en la dirección del gradiente de temperatura en el punto (1, 1).

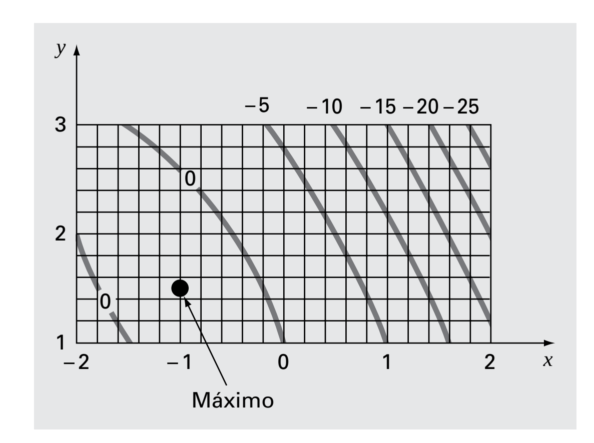


Figura P14.10
La búsqueda por malla.

CAPÍTULO 15

Optimización con restricciones

Este capítulo aborda problemas de optimización en los cuales entran en juego restricciones. Primero, se analizarán problemas donde la función objetivo y las restricciones son lineales. Para tales casos, hay métodos especiales que aprovechan la linealidad de las funciones, llamados métodos de programación lineal. Los algoritmos resultantes resuelven con gran eficiencia problemas muy grandes con miles de variables y restricciones. Dichos métodos se utilizan en una gran variedad de problemas en ingeniería y en administración.

Después, se verá en forma breve el problema más general de optimización con restricciones no lineal. Finalmente, se proporcionará una visión general de cómo se emplean los paquetes de software en la optimización.

15.1 PROGRAMACIÓN LINEAL

La *programación lineal* (o PL, por simplicidad) es un método de optimización que se ocupa del cumplimiento de un determinado objetivo, como maximizar las utilidades o minimizar el costo, en presencia de restricciones como recursos limitados. El término *lineal* denota que las funciones matemáticas que representan el objetivo y las restricciones son lineales. El término *programación* no significa "programación en computadora"; más bien denota "programar" o "fijar una agenda" (Revelle y colaboradores, 1997).

15.1.1 Forma estándar

El problema básico de la programación lineal consiste en dos partes principales: la función objetivo y un conjunto de restricciones. En un problema de maximización, la función objetivo, por lo general, se expresa como

$$\text{Maximizar } Z = c_1x_1 + c_2x_2 + \cdots + c_nx_n \tag{15.1}$$

donde c_j = la contribución de cada unidad de la j-ésima actividad realizada y x_j = magnitud de la j-ésima actividad. Así, el valor de la función objetivo, Z, es la contribución total debida al número total de actividades, n.

Las restricciones se representan, en forma general, como

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \le b_i \tag{15.2}$$

donde a_{ij} = cantidad del i-ésimo recurso que se consume por cada unidad de la j-ésima actividad, y b_i = cantidad del i-ésimo recurso que está disponible. Es decir, los recursos son limitados.

El segundo tipo general de restricción, especifica que todas las actividades deben tener un valor positivo:

$$x_i \ge 0 \tag{15.3}$$

En el presente contexto, lo anterior expresa la noción realista de que, en algunos problemas, la actividad negativa es físicamente imposible (por ejemplo, no se pueden producir bienes negativos).

Juntas, la función objetivo y las restricciones, especifican el problema de programación lineal. Éstas indican que se trata de maximizar la contribución de varias actividades, bajo la restricción de que en estas actividades se utilizan cantidades finitas de recursos. Antes de mostrar cómo se puede obtener este resultado, primero se desarrollará un ejemplo.

EJEMPLO 15.1 Planteamiento del problema de la PL

Planteamiento del problema Se desarrolla el siguiente problema del área de la ingeniería química o petrolera. Aunque, éste es relevante para todas las áreas de la ingeniería relacionadas con la generación de productos con recursos limitados.

Suponga que una planta procesadora de gas recibe cada semana una cantidad fija de materia prima. Esta última se procesa para dar dos tipos de gas: calidad regular y prémium. Estas clases de gas son de gran demanda (es decir, se tiene garantizada su venta) y dan diferentes utilidades a la compañía. Sin embargo, su producción involucra restricciones de tiempo y de almacenamiento. Por ejemplo, no se pueden producir las dos clases a la vez, y las instalaciones están disponibles solamente 80 horas por semana. Además, existe un límite de almacenamiento para cada uno de los productos. Todos estos factores se enlistan abajo (observe que una tonelada métrica, o *ton*, es igual a 1 000 kg):

	Producto		
Recurso	**Regular**	**Prémium**	**Disponibilidad del recurso**
Materia prima	7 m^3/ton	11 m^3/ton	77 m^3/semana
Tiempo de producción	10 h/ton	8 h/ton	80 h/semana
Almacenamiento	9 ton	6 ton	
Ganancia	150/ton	175/ton	

Desarrolle una formulación de programación lineal para maximizar las utilidades de esta operación.

Solución El ingeniero que opera esta planta debe decidir la cantidad a producir de cada tipo de gas para maximizar las utilidades. Si las cantidades producidas cada semana de gas regular y prémium se designan como x_1 y x_2, respectivamente, la ganancia total se calcula mediante

$$\text{Ganancia total} = 150x_1 + 175x_2$$

o se escribe como una función objetivo en programación lineal:

$$\text{Maximizar } Z = 150x_1 + 175x_2$$

Las restricciones se desarrollan en una forma similar. Por ejemplo, el total de gas crudo (materia prima) utilizado se calcula como:

$$\text{Total de gas utilizado} = 7x_1 + 11x_2$$

Este total no puede exceder el abastecimiento disponible de 77 m^3/semana, así que la restricción se representa como

$$7x_1 + 11x_2 \le 77$$

Las restricciones restantes se desarrollan en una forma similar: la formulación completa resultante de PL está dada por

$$\text{Maximizar } Z = 150x_1 + 175x_2 \qquad \text{(maximizar la ganancia)}$$

sujeta a

$$7x_1 + 11x_2 \le 77 \qquad \text{(restricciones de material)}$$
$$10x_1 + 8x_2 \le 80 \qquad \text{(restricción de tiempo)}$$
$$x_1 \le 9 \qquad \text{(restricción de almacenaje de gas “regular”)}$$
$$x_2 \le 6 \qquad \text{(restricción de almacenaje de gas “prémium”)}$$
$$x_1, x_2 \ge 0 \qquad \text{(restricciones positivas)}$$

Observe que el conjunto de ecuaciones anterior constituye la formulación completa de PL. Las explicaciones en los paréntesis de la derecha se han incluido para aclarar el significado de cada expresión.

15.1.2 Solución gráfica

Debido a que las soluciones gráficas están limitadas a dos o tres dimensiones, tienen una utilidad práctica limitada. Sin embargo, son muy útiles para demostrar algunos conceptos básicos de las técnicas algebraicas generales, utilizadas para resolver problemas multidimensiones en la computadora.

En un problema bidimensional, como el del ejemplo 15.1, el espacio solución se define como un plano con x_1 medida a lo largo de la abscisa; y x_2, a lo largo de la ordenada. Como las restricciones son lineales, se trazan sobre este plano como líneas rectas. Si el problema de PL se formuló adecuadamente (es decir, si tiene una solución), estas líneas restrictivas describen una región, llamada el *espacio de solución factible*, que abarca todas las posibles combinaciones de x_1 y x_2, las cuales obedecen las restricciones y, por lo tanto, representan soluciones factibles. La función objetivo de un valor particular de Z se puede trazar como otra línea recta y sobreponerse en este espacio. El valor de Z, entonces, se ajusta hasta que esté en el valor máximo, y toque aún el espacio factible. Este valor de Z representa la solución óptima. Los valores correspondientes de x_1 y x_2, donde Z toca el espacio de solución factible, representan los valores óptimos de las actividades. El siguiente ejemplo deberá ayudar a aclarar el procedimiento.

EJEMPLO 15.2 Solución gráfica

Planteamiento del problema Desarrolle una solución gráfica para el problema del procesamiento de gas del ejemplo 15.1:

$$\text{Maximizar } Z = 150x_1 + 175x_2$$

sujeta a

$$7x_1 + 11x_2 \leq 77 \quad (1)$$

$$10x_1 + 8x_2 \leq 80 \quad (2)$$

$$x_1 \leq 9 \quad (3)$$

$$x_2 \leq 6 \quad (4)$$

$$x_1 \geq 0 \quad (5)$$

$$x_2 \geq 0 \quad (6)$$

Se numeraron las restricciones para identificarlas en la siguiente solución gráfica.

Solución Primero, se trazan las restricciones sobre el espacio solución. Por ejemplo, se reformula la primera restricción como una línea al reemplazar la desigualdad por un signo igual, y se despeja x_2:

$$x_2 = -\frac{7}{11}x_1 + 7$$

Así, como en la figura 15.1*a*, los valores posibles de x_1 y x_2 que obedecen dicha restricción se hallan por debajo de esta línea (en la gráfica, la dirección se indica con la pequeña flecha). Las otras restricciones se evalúan en forma similar, se sobreponen en la figura 15.1*a*. Observe cómo éstas encierran una región donde todas se satisfacen. Éste es el espacio solución factible (el área *ABCDE* en la gráfica).

Además de definir el espacio factible, la figura 15.1*a* también ofrece una mejor comprensión. En particular, se percibe que la restricción 3 (almacenamiento de gas regular) es “redundante”. Es decir, el espacio solución factible no resulta afectado si fuese suprimida.

Después, se agrega la función objetivo a la gráfica. Para hacerlo, se debe escoger un valor de Z. Por ejemplo, para $Z = 0$, la función objetivo es ahora

$$0 = 150x_1 + 175x_2$$

o, despejando x_2, se obtiene la línea recta

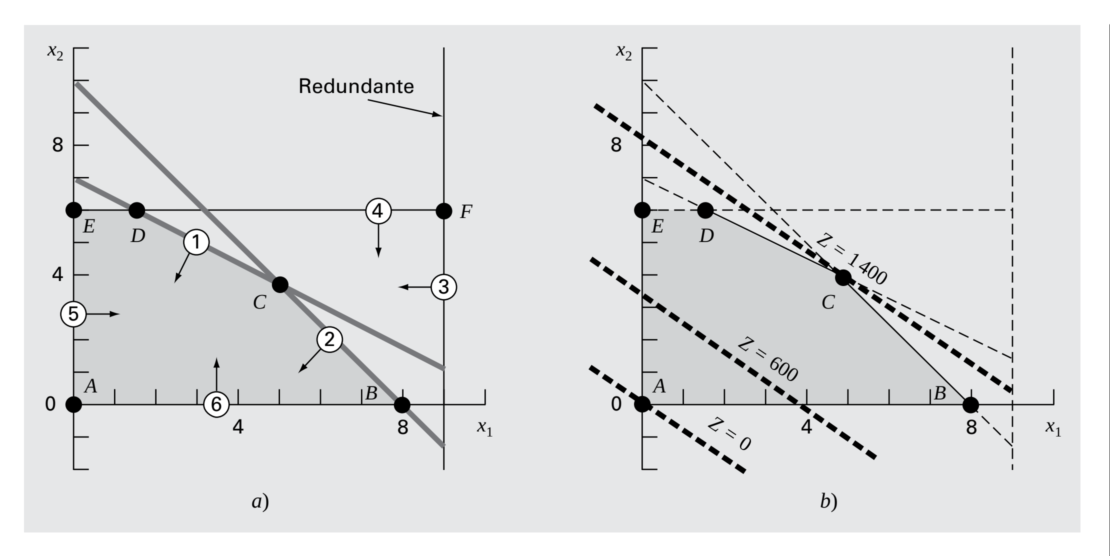


Figura 15.1
Solución gráfica de un problema de programación lineal. *a*) Las restricciones definen un espacio solución factible. *b*) La función objetivo se incrementa hasta que alcance el valor máximo que cumpla con todas las restricciones. Gráficamente, se mueve hacia arriba y a la derecha, hasta que toca el espacio factible en un solo punto óptimo.

$$x_2 = -\frac{150}{175}x_1$$

Como se muestra en la figura 15.1*b*, ésta representa una línea punteada que interseca el origen. Ahora, debido a que estamos interesados en maximizar *Z*, ésta se aumenta a, digamos, 600, y la función objetivo es

$$x_2 = \frac{600}{175} - \frac{150}{175}x_1$$

Así, al incrementar el valor de la función objetivo, la línea se aleja del origen. Como la línea todavía cae dentro del espacio solución, nuestro resultado es aún factible. No obstante, por la misma razón, todavía hay espacio para mejorarlo. Por lo tanto, *Z* continúa aumentando hasta que un incremento adicional lleve la función objetivo más allá de la región factible. Como se muestra en la figura 15.1*b*, el valor máximo de *Z* corresponde aproximadamente a 1 400. En este punto, x_1 y x_2 son, de manera aproximada, iguales a 4.9 y 3.9, respectivamente. Así, la solución gráfica indica que si se producen estas cantidades de gas regular y prémium, se alcanzará una máxima utilidad de aproximadamente 1 400.

Además de determinar los valores óptimos, el procedimiento gráfico ofrece una mejor comprensión del problema. Esto se aprecia al sustituir de nuevo las soluciones en las ecuaciones restrictivas:

$$7(4.9) + 11(3.9) \cong 77$$
$$10(4.9) + 8(3.9) \cong 80$$
$$4.9 \leq 9$$
$$3.9 \leq 6$$

En consecuencia, como se ve claramente en la gráfica, producir la cantidad óptima de cada producto nos lleva directamente al punto donde se satisfacen las restricciones de los recursos (1) y del tiempo (2). Tales restricciones se dice que son *obligatorias*. Además, la gráfica también hace evidente que ninguna de las restricciones de almacenamiento [(3) ni (4)] actúan como una limitante. Tales restricciones se conocen como *no obligatorias*. Esto nos lleva a la conclusión práctica

de que, en este caso, se pueden aumentar las utilidades, ya sea con un incremento en el abastecimiento de recursos (el gas crudo) o en tiempo de producción. Además, esto indica que el aumento del almacenamiento podría no tener impacto sobre las utilidades.

El resultado obtenido en el ejemplo anterior es uno de los cuatro posibles resultados que, por lo general, se obtienen en un problema de programación lineal. Éstos son:

1. *Solución única.* Como en el ejemplo, la función objetivo máxima interseca un solo punto.
2. *Soluciones alternativas.* Suponga que los coeficientes de la función objetivo del ejemplo fueran paralelos precisamente a una de las restricciones. En nuestro ejemplo, una forma en la cual esto podría ocurrir sería que las utilidades se modificaran a $140/ton y $220/ton. Entonces, en lugar de un solo punto, el problema podría tener un número infinito de óptimos correspondientes a un segmento de línea (véase figura 15.2*a*).

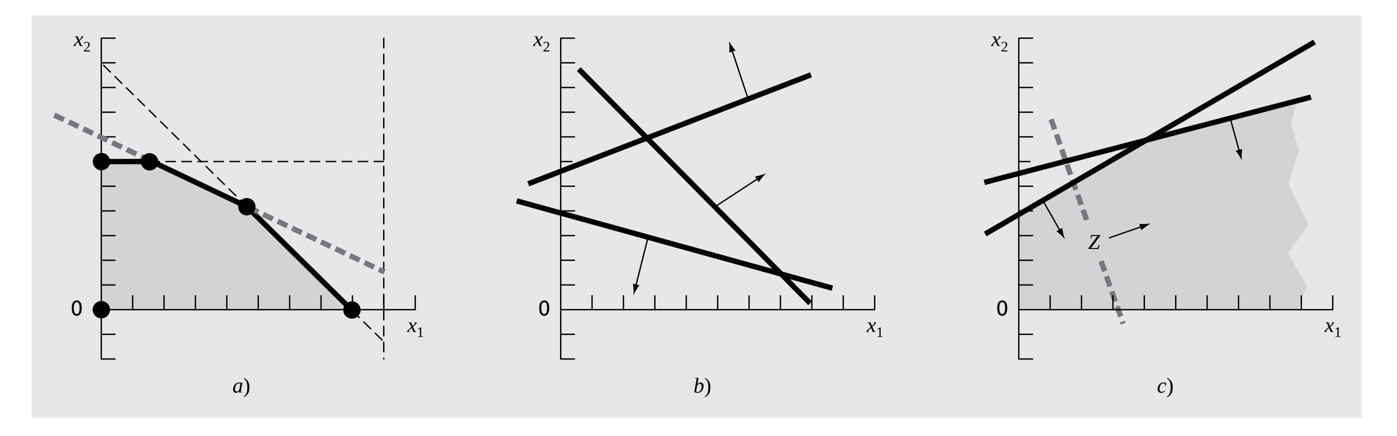


Figura 15.2
Además de una sola solución óptima (por ejemplo, la figura 15.1*b*), existen otros tres resultados posibles en un problema de programación lineal: *a*) óptima alternativa, *b*) solución no factible y *c*) un resultado no acotado.

3. *Solución no factible.* Como en la figura 15.2*b*, es posible que el problema esté formulado de tal manera que no haya una solución factible. Esto puede deberse a que se trata de un problema sin solución o a errores en la formulación del problema. Lo último ocurre si el problema está sobrerrestringido, y ninguna solución satisface todas las restricciones.
4. *Problemas no acotados.* Como en la figura 15.2*c*, esto usualmente significa que el problema está subrestringido y, por lo tanto, tiene límites abiertos. Como en el caso de la solución no factible, esto a menudo ocurre debido a errores cometidos durante la especificación del problema.

Ahora supongamos que nuestro problema tiene una solución única. El procedimiento gráfico podría sugerir una estrategia numérica para dar con el máximo. Observando la figura 15.1, deberá quedar claro que siempre se presenta el óptimo en uno de los puntos esquina, donde se presentan dos restricciones. Tales puntos se conocen de manera formal como *puntos extremos*. Así, del número infinito de posibilidades en el espacio de decisión, al enfocarse en los puntos extremos, se reducen claramente las opciones posibles.

Además, es posible reconocer que no todo punto extremo es factible; es decir, satisface todas las restricciones. Por ejemplo, observe que el punto F en la figura 15.1*a* es un punto extremo; pero no es factible. Limitándonos a *puntos extremos factibles*, se reduce todavía más el campo factible.

Por último, una vez que se han identificado todos los puntos extremos factibles, el que ofrezca el mejor valor de la función objetivo representará la solución óptima. Se podría encontrar esta solución óptima mediante la exhaustiva (e ineficiente) evaluación del valor de la función objetivo en cada punto extremo factible. En la siguiente sección se analiza el método simplex, que ofrece una mejor estrategia para trazar un rumbo selectivo, a través de una secuencia de puntos extremos factibles, para llegar al óptimo de una manera extremadamente eficiente.

15.1.3 El método simplex

El método simplex se basa en la suposición de que la solución óptima estará en un punto extremo. Así, el procedimiento debe ser capaz de discriminar si durante la solución del problema se presentará un punto extremo. Para esto, las ecuaciones con restricciones se reformulan como igualdades, introduciendo las llamadas variables de holgura.

Variables de holgura Como lo indica su nombre, una *variable de holgura* mide cuánto de un recurso restringido está disponible; es decir, cuánta "holgura" está disponible. Por ejemplo, recuerde el recurso restringido que se utilizó en los ejemplos 15.1 y 15.2:

$$7x_1 + 11x_2 \leq 77$$

Se define una variable de holgura S_1 como la cantidad de gas crudo que no se usa para un nivel de producción específico (x_1, x_2). Si esta cantidad se suma al lado izquierdo de la restricción, esto vuelve exacta a la relación:

$$7x_1 + 11x_2 + S_1 = 77$$

Ahora vemos qué nos dice la variable de holgura. Si es positiva, significa que se tiene algo de "holgura" en esta restricción. Es decir, se cuenta con un excedente de recurso que no se está utilizando por completo. Si es negativa, nos indica que hemos sobrepasado la restricción. Finalmente, si es cero, denota que la restricción se satisface con precisión. Es decir, hemos utilizado todo el recurso disponible. Puesto que ésta es exactamente la condición donde las líneas de restricción se intersecan, la variable de holgura ofrece un medio para detectar los puntos extremos.

Una variable de holgura diferente se desarrolla para cada ecuación restringida, lo cual resulta en lo que se conoce como la *versión aumentada completamente*,

$$\text{Maximizar } Z = 150x_1 + 175x_2$$

sujeta a

$$7x_1 + 11x_2 + S_1 = 77 \qquad (15.4a)$$

$$10x_1 + 8x_2 + S_2 = 80 \qquad (15.4b)$$

$$x_1 + S_3 = 9 \qquad (15.4c)$$

$$x_2 + S_4 = 6 \qquad (15.4d)$$

$$x_1, x_2, S_1, S_2, S_3, S_4 \geq 0$$

Advierta cómo se han establecido las cuatro ecuaciones, de manera que las incógnitas quedan alineadas en columnas. Se hizo así para resaltar que ahora se trata de un sistema de ecuaciones algebraicas lineales (recuerde la parte tres). En la siguiente sección se mostrará cómo se emplean dichas ecuaciones para determinar los puntos extremos en forma algebraica.

Solución algebraica A diferencia de la parte tres, donde se tenían n ecuaciones con n incógnitas, nuestro sistema del ejemplo [ecuaciones (15.4)] está *subespecificado* o *indeterminado*; es decir, tiene más incógnitas que ecuaciones. En términos generales, hay *n variables estructurales* (las incógnitas originales), *m variables de holgura* o *excedentes* (una por restricción) y $n + m$ variables en total (estructurales más excedentes). En el problema de la producción de gas se tienen 2 variables estructurales, 4 variables de holgura y 6 variables en total. Así, el problema consiste en resolver 4 ecuaciones con 6 incógnitas.

La diferencia entre el número de incógnitas y el de ecuaciones (igual a 2 en nuestro problema) está directamente relacionada con la forma en que se distingue un punto extremo factible. Específicamente, cada punto factible tiene 2 de las 6 variables igualadas a cero. Por ejemplo, los cinco puntos en las esquinas del área *ABCDE* tienen los siguientes valores cero:

Punto extremo	**Variables cero**
A	x_1, x_2
B	x_2, S_2
C	S_1, S_2
D	S_1, S_4
E	x_1, S_4

Esta observación nos lleva a concluir que los puntos extremos se determinan a partir de la forma estándar igualando dos de las variables a cero. En nuestro ejemplo, esto reduce el problema a resolver 4 ecuaciones con 4 incógnitas. Por ejemplo, para el punto E, si $x_1 = S_4 = 0$, la forma estándar se reduce a

$$\begin{aligned} 11x_2 + S_1 \qquad\qquad &= 77 \\ 8x_2 \qquad + S_2 \qquad &= 80 \\ + S_3 &= 9 \\ x_2 \qquad\qquad\qquad &= 6 \end{aligned}$$

de donde se obtiene $x_2 = 6$, $S_1 = 11$, $S_2 = 32$ y $S_3 = 9$. Junto con $x_1 = S_4 = 0$, estos valores definen el punto E.

Generalizando, una solución básica de m ecuaciones lineales con n incógnitas se obtiene al igualar a cero las variables $n - m$ y resolver las m ecuaciones para las m incógnitas restantes. Las variables igualadas a cero se conocen formalmente como *variables no básicas*; mientras que a las m variables restantes se les llama *variables básicas*. Si todas las variables básicas son no negativas, al resultado se le llama una *solución factible básica*. El óptimo será una de éstas.

Ahora, un procedimiento directo para determinar la solución óptima será calcular todas las soluciones básicas, determinar cuáles de ellas son factibles, y de éstas, cuál tiene el valor mayor de Z. Sin embargo, éste no es un procedimiento recomendable por dos razones.

Primero, aun para problemas de tamaño moderado, se necesita resolver una gran cantidad de ecuaciones. Para m ecuaciones con n incógnitas, se tendrán que resolver

$$C_m^n = \frac{n!}{m!(n-m)!}$$

ecuaciones simultáneas. Por ejemplo, si hay 10 ecuaciones ($m = 10$) con 16 incógnitas ($n = 16$), ¡se tendrían 8 008 [= 16!/(10!6!)] sistemas de ecuaciones de 10×10 para resolver!

Segundo, quizás una porción significativa de éstas no sea factible. Por ejemplo, en el problema actual de los $C_6^4 = 15$ puntos extremos, sólo 5 son factibles. Claramente, si se pudiese evitar resolver todos estos sistemas innecesarios, se tendría un algoritmo más eficiente. Uno de estos procedimientos se describe a continuación.

Implementación del método simplex El método simplex evita las ineficiencias descritas en la sección anterior. Esto se hace al comenzar con una solución factible básica. Luego se mueve a través de una secuencia de otras soluciones factibles básicas que mejoran sucesivamente el valor de la función objetivo. En forma eventual, se alcanza el valor óptimo y se termina el método.

Se ilustrará el procedimiento con el problema de procesamiento de gas, de los ejemplos 15.1 y 15.2. El primer paso consiste en empezar en una solución factible básica (es decir, en un punto esquina extremo del espacio factible). Para casos como los nuestros, un punto de inicio obvio podría ser el punto A; esto es, $x_1 = x_2 = 0$. Las 6 ecuaciones originales en 4 incógnitas se convierten en

$$\begin{aligned} S_1 \qquad\qquad\qquad &= 77 \\ S_2 \qquad\qquad &= 80 \\ S_3 \qquad &= 9 \\ S_4 &= 6 \end{aligned}$$

Así, los valores iniciales de las variables básicas se dan automáticamente y son iguales a los lados derecho de las restricciones.

Antes de proceder al siguiente paso, la información inicial se puede resumir en un adecuado formato tabular. Como se muestra a continuación, la *tabla* proporciona un resumen de la información clave que constituye el problema de la programación lineal.

Básica	Z	x_1	x_2	S_1	S_2	S_3	S_4	**Solución**	**Intersección**
Z	1	–150	–175	0	0	0	0	0	
S_1	0	7	11	1	0	0	0	77	11
S_2	0	10	8	0	1	0	0	80	8
S_3	0	1	0	0	0	1	0	9	9
S_4	0	0	1	0	0	0	1	6	∞

Observe que para propósitos de la tabla, la función objetivo se expresa como

$$Z - 150x_1 - 175x_2 - 0S_1 - 0S_2 - 0S_3 - 0S_4 = 0 \tag{15.5}$$

El siguiente paso consiste en moverse a una nueva solución factible básica que nos lleve a mejorar la función objetivo. Esto se consigue incrementando una variable actual no básica (en este punto, x_1 o x_2) por arriba de cero para que Z aumente. Recuerde que, en el ejemplo presente, los puntos extremos deben tener 2 valores cero. Por lo tanto, una de las variables básicas actuales (S_1, S_2, S_3 o S_4) también deben igualarse a cero.

Para resumir este paso importante: una de las variables no básicas actuales debe hacerse básica (no cero). Esta variable se llama *variable de entrada*. En el proceso, una de las variables básicas actuales se vuelve no básica (cero). Esta variable se llama *variable de salida*.

Ahora, desarrollaremos un procedimiento matemático para seleccionar las variables de entrada y de salida. A causa de la convención de cómo escribir la función objetivo [ecuación (15.5)], la variable de entrada puede ser cualquier variable de la función objetivo que tenga un coeficiente negativo (ya que esto hará a Z más grande). La variable con el valor negativo más grande se elige de manera convencional porque usualmente nos lleva al incremento mayor en Z. En nuestro caso, x_2 será la variable entrante puesto que su coeficiente, –175, es más negativo que el coeficiente de x_1: –150.

Aquí se puede consultar la solución gráfica para mejor comprensión. Se comienza en el punto inicial A, como se muestra en la figura 15.3. Considerando su coeficiente, se escogerá x_2 como entrada. No obstante, para abreviar en este ejemplo, seleccionamos x_1 puesto que en la gráfica se observa que nos llevará más rápido al máximo.

Después, se debe elegir la variable de salida entre las variables básicas actuales (S_1, S_2, S_3 o S_4). Se observa gráficamente que hay dos posibilidades. Moviéndonos al punto B se tendrá S_2 igual a cero; mientras que al movernos al punto F tendremos S_1 igual a cero. Sin embargo, en la gráfica también queda claro que F no es posible, ya que queda fuera del espacio solución factible. Así, decide moverse de A a B.

¿Cómo se detecta el mismo resultado en forma matemática? Una manera es calcular los valores en los que las líneas de restricción intersecan el eje o la línea que corresponde a la variable saliente (en nuestro caso, el eje x_1). Es posible calcular este valor como la razón del lado derecho de la restricción (la columna "Solución" de la tabla) entre el coeficiente correspondiente de x_1. Por ejemplo, para la primera variable de holgura restrictiva S_1, el resultado es

$$\text{Intersección} = \frac{77}{7} = 11$$

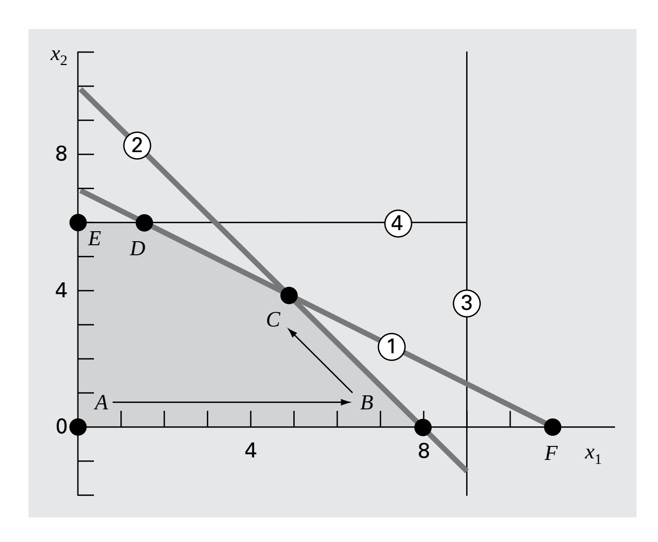


Figura 15.3
Ilustración gráfica de cómo se mueve en forma sucesiva el método simplex a través de soluciones básicas factibles, para llegar al óptimo de manera eficiente.

Las intersecciones restantes se pueden calcular y enlistar como la última columna de la tabla. Debido a que 8 es la menor intersección positiva, significa que la segunda línea de restricción se alcanzará primero conforme se incremente x_1. Por lo tanto, S_2 será la variable de entrada.

De esta manera, nos hemos movido al punto B ($x_2 = S_2 = 0$), y la nueva solución básica es ahora

$$
\begin{aligned}
7x_1 + S_1 \qquad &= 77 \\
10x_1 \qquad &= 80 \\
x_1 \qquad + S_3 \quad &= 9 \\
S_4 &= 6
\end{aligned}
$$

La solución de este sistema de ecuaciones define efectivamente los valores de las variables básicas en el punto B: $x_1 = 8$, $S_1 = 21$, $S_3 = 1$ y $S_4 = 6$.

Se utiliza la tabla para realizar los mismos cálculos empleando el método de Gauss-Jordan. Recuerde que la estrategia básica de este método implica convertir el elemento pivote en 1, y después eliminar los coeficientes en la misma columna arriba y abajo del elemento pivote (recuerde la sección 9.7).

En este ejemplo, el renglón pivote es S_2 (la variable de entrada) y el elemento pivote es 10 (el coeficiente de la variable de salida, x_1). Al dividir el renglón entre 10 y reemplazar S_2 por x_1 se tiene

Básica	Z	x_1	x_2	S_1	S_2	S_3	S_4	**Solución**	**Intersección**
Z	1	–150	–175	0	0	0	0	0	
S_1	0	7	11	1	0	0	0	77	
x_1	0	1	0.8	0	0.1	0	0	8	
S_3	0	1	0	0	0	1	0	9	
S_4	0	0	1	0	0	0	1	6	

Después, se eliminan los coeficientes de x_1 en los otros renglones. Por ejemplo, para el renglón de la función objetivo, el renglón pivote se multiplica por –150 y el resultado se resta del primer renglón para obtener

Z	x_1	x_2	S_1	S_2	S_3	S_4	**Solución**
1	–150	–175	0	0	0	0	0
–0	–(–150)	–(–120)	–0	–(–15)	0	0	–(–1 200)
1	0	–55	0	15	0	0	1 200

Es posible realizar operaciones similares en los renglones restantes para obtener la nueva tabla,

Básica	Z	x_1	x_2	S_1	S_2	S_3	S_4	**Solución**	**Intersección**
Z	1	0	–55	0	15	0	0	1 200	
S_1	0	0	5.4	1	–0.7	0	0	21	3.889
x_1	0	1	0.8	0	0.1	0	0	8	10
S_3	0	0	–0.8	0	–0.1	1	0	1	–1.25
S_4	0	0	1	0	0	0	1	6	6

Así la nueva tabla resume toda la información del punto B. Esto incluye el hecho de que el movimiento ha aumentado la función objetivo a $Z = 1\,200$.

Esta tabla se utiliza después para representar el próximo y, en este caso, último paso. Sólo una variable más, x_2, tiene un valor negativo en la función objetivo, y se elige, por lo tanto, como la variable de salida. De acuerdo con los valores de la intersección (ahora calculados como la columna solución sobre los coeficientes de la columna de x_2), la primera restricción tiene el valor positivo más

pequeño y, por lo tanto, se selecciona S_1 como la variable de entrada. Así, el método simplex nos mueve del punto *B* al *C* en la figura 15.3. Por último, la eliminación de Gauss-Jordan se utiliza para resolver las ecuaciones simultáneas. El resultado es la tabla final,

Básica	Z	x_1	x_2	S_1	S_2	S_3	S_4	Solución
Z	1	0	0	10.1852	7.8704	0	0	1 413.889
x_2	0	0	1	0.1852	–0.1296	0	0	3.889
x_1	0	1	0	–0.1481	0.2037	0	0	4.889
S_3	0	0	0	0.1481	–0.2037	1	0	4.111
S_4	0	0	0	–0.1852	0.1296	0	1	2.111

Se sabe que éste es el resultado final porque no quedan coeficientes negativos en la fila de la función objetivo. La solución final se tabula como $x_1 = 3.889$ y $x_2 = 4.889$, que dan una función objetivo máxima $Z = 1\,413.889$. Además, como S_3 y S_4 están todavía en la base, sabemos que la solución está limitada por la primera y la segunda restricciones.

15.2 OPTIMIZACIÓN NO LINEAL CON RESTRICCIONES

Existen varios procedimientos para los problemas de optimización no lineal con la presencia de restricciones. Generalmente, dichos procedimientos se dividen en directos e indirectos (Rao, 1996). Los procedimientos indirectos típicos usan las llamadas *funciones de penalización*. Éstas consideran expresiones adicionales para hacer que la función objetivo sea menos óptima conforme la solución se aproxima a una restricción. Así, la solución no será aceptada por violar las restricciones. Aunque tales métodos llegan a ser útiles en algunos problemas, se vuelven difíciles cuando el problema tiene muchas restricciones.

El método de búsqueda del *gradiente reducido generalizado*, o GRG, es uno de los métodos directos más populares (para detalles, véase Fylstra *et al*., 1998; Lasdon *et al*., 1978; Lasdon y Smith, 1992). Éste es, de hecho, el método no lineal usado en el Solver de Excel.

Este método primero "reduce" a un problema de optimización sin restricciones. Lo hace resolviendo en un conjunto de ecuaciones no lineales las variables básicas en términos de variables no básicas. Después, se resuelve el problema sin restricciones utilizando procedimientos similares a los que se describen en el capítulo 14. Se escoge primero una dirección de búsqueda a lo largo de la cual se busca mejorar la función objetivo. La selección obvia es un *procedimiento cuasi-Newton* (BFGS) que, como se describió en el capítulo 14, requiere el almacenamiento de una aproximación de la matriz hessiana. Este procedimiento funciona muy bien en la mayoría de los casos. El procedimiento del *gradiente conjugado* también está disponible en Excel como una alternativa para problemas grandes. El Solver de Excel tiene la excelente característica de que, en forma automática, cambia al método del gradiente conjugado, dependiendo de la capacidad de almacenamiento. Una vez establecida la dirección de búsqueda, se lleva a cabo una búsqueda unidimensional a lo largo de esa dirección, mediante un procedimiento de tamaño de paso variable.

15.3 OPTIMIZACIÓN CON PAQUETES DE SOFTWARE

Los paquetes de software tienen grandes capacidades para la optimización. En esta sección, se dará una introducción a algunos de los más útiles.

15.3.1 Programación lineal en Excel

Existe una variedad de paquetes de software especialmente diseñados para la programación lineal. Sin embargo, como su disponibilidad es amplia, este análisis se concentrará en la hoja de cálculo de Excel. Ésta usa la opción Solver que se estudió, previamente en el capítulo 7, para localizar raíces.

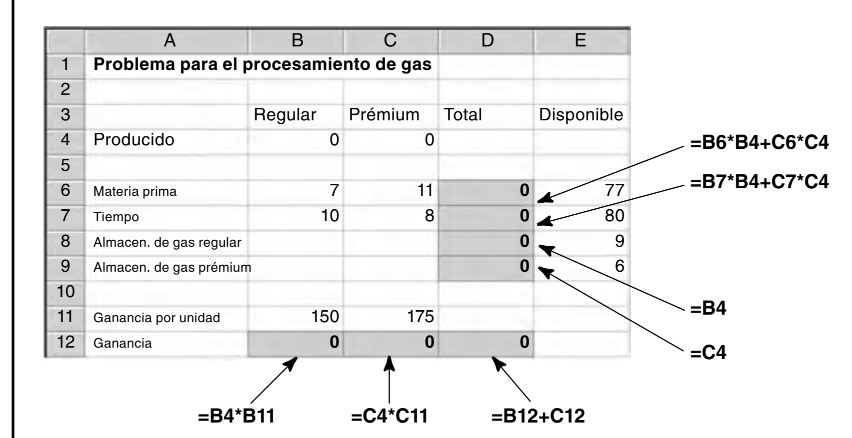
La manera en la cual se usa Solver para programación lineal es similar a las aplicaciones hechas con anterioridad, en el sentido de cómo se introducen los datos en las celdas de la hoja de cálculo. La estrategia básica consiste en ubicar una celda que será optimizada, como una función de las va-

riaciones de las otras celdas sobre la misma hoja de cálculo. El siguiente ejemplo ilustra cómo realizar esto con el problema del procesamiento de gas.

EJEMPLO 15.3 Uso del Solver de Excel para un problema de programación lineal

Planteamiento del problema Utilice Excel para resolver el problema del procesamiento de gas que examinamos en este capítulo.

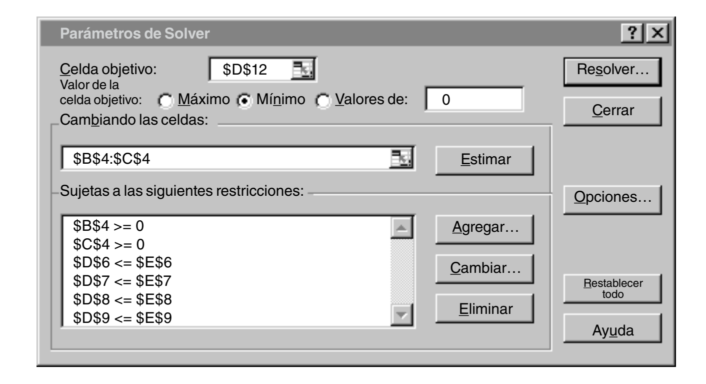
Solución Una hoja de cálculo de Excel para calcular los valores pertinentes en el problema del procesamiento de gas se muestra en la figura 15.4. Las celdas no sombreadas son las que contienen los datos numéricos y las leyendas. Las celdas sombreadas contienen las cantidades que se calculan basadas en las otras celdas. Reconozca que la celda que va a ser maximizada es la D12, la cual contiene la utilidad total. Las celdas que cambian son B4:C4, donde se tienen las cantidades producidas de gas regular y prémium.



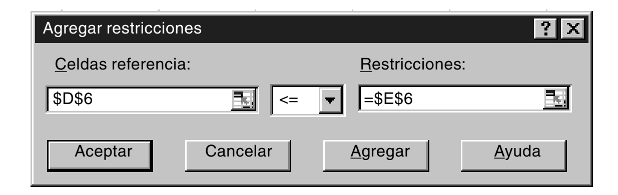
	A	B	C	D	E
1	**Problema para el procesamiento de gas**				
2					
3		Regular	Prémium	Total	Disponible
4	Producido	0	0		
5					
6	Materia prima	7	11	**0**	77
7	Tiempo	10	8	**0**	80
8	Almacen. de gas regular			**0**	9
9	Almacen. de gas prémium			**0**	6
10					
11	Ganancia por unidad	150	175		
12	Ganancia	**0**	**0**	**0**	

Figura 15.4
Hoja de cálculo en Excel para usar el Solver con la programación lineal.

Una vez que se crea la hoja de cálculo, se selecciona **Solver** del menú Tools (Herramientas). En este momento se despliega una ventana de diálogo solicitándole la información pertinente. Las celdas del cuadro de diálogo de Solver se llenan así



Las restricciones se agregan una por una al seleccionar el botón "Agregar". Esto abrirá un cuadro de diálogo que se ve así



Entonces, la restricción donde el total de materia prima (celda D6) debe ser menor o igual que el abastecimiento disponible (E6) se agrega como se muestra. Después de agregar cada una de las restricciones, puede seleccionarse el botón agregar. Cuando se hayan introducido las cuatro restricciones, seleccionamos el botón Aceptar para regresar a la ventana de diálogo del Solver.

Ahora, antes de la ejecución, se deberá seleccionar el botón "Opciones…" del Solver y escoger el cuadro rotulado como "Assume linear model" (Suponer modelo lineal). Esto hará que Excel emplee una versión del algoritmo simplex (en lugar del Solver no lineal más general que normalmente usa) que acelera su aplicación.

Después de seleccionar esta opción, regrese el menú Solver. Cuando seleccione el botón aceptar, se abrirá un cuadro de diálogo con un reporte sobre el éxito de la operación. En el caso actual, el Solver obtiene la solución correcta (figura 15.5).

	A	B	C	D	E
1	**Problema para el procesamiento de gas**				
2					
3		Regular	Prémium	Total	Disponible
4	Producido	**4.888889**	**3.888889**		
5					
6	Materia prima	7	11	77	77
7	Tiempo	10	8	80	80
8	Almacen. de gas regular			4.888889	9
9	Almacen. de gas prémium			3.888889	6
10					
11	Ganancia por unidad	150	175		
12	Ganancia	733.3333	680.5556	**1413.889**	

Figura 15.5
Hoja de cálculo de Excel con la solución al problema de programación lineal.

Además de obtener la solución, Solver también ofrece algunos reportes en resumen útiles. Éstos se explorarán en las aplicaciones a la ingeniería que se describen en la sección 16.2.

15.3.2 Excel para la optimización no lineal

La manera de usar Solver para la optimización no lineal es similar a las aplicaciones hechas con anterioridad en el sentido de cómo los datos se introducen en las celdas de la hoja de cálculo. Una vez más, la estrategia básica es tener una sola celda a optimizar, como una función de las variaciones en las otras celdas sobre la misma hoja de cálculo. El siguiente ejemplo ilustra cómo hacer esto con el problema del paracaidista que planteamos en la introducción de esta parte del libro (recuerde el ejemplo PT4.1).

EJEMPLO 15.4 Uso del Solver de Excel para la optimización con restricciones no lineal

Planteamiento del problema Recuerde que en el ejemplo PT4.1 se desarrolló una optimización con restricciones no lineal para minimizar el costo de la caída de un paracaídas en un campo de refugiados. Los parámetros de este problema son

Parámetro	Símbolo	Valor	Unidades
Masa total	M_t	2 000	kg
Aceleración de la gravedad	g	9.8	m/s^2
Coeficiente de costo (constante)	c_0	200	\$
Coeficiente de costo (longitud)	c_1	56	\$/m
Coeficiente de costo (área)	c_2	0.1	\$/m^2
Velocidad crítica de impacto	v_c	20	m/s
Efecto del área sobre el arrastre	k_c	3	kg/(s · m^2)
Altura inicial de caída	z_0	500	m

Sustituyendo estos valores en las ecuaciones (PT4.11) a (PT4.19) se obtiene

Minimizar $C = n(200 + 56\ell + 0.1A^2)$

sujeta a

$$v \leq 20$$
$$n \geq 1$$

donde n es un entero y todas las otras variables son reales. Además, las siguientes cantidades se definen como

$$A = 2\pi r^2$$
$$\ell = \sqrt{2}\, r$$
$$c = 3A$$
$$m = \frac{M_t}{n} \tag{E15.4.1}$$
$$t = \text{raíz}\left[500 - \frac{9.8m}{c}t + \frac{9.8m^2}{c^2}(1 - e^{-(c/m)t})\right] \tag{E15.4.2}$$
$$v = \frac{9.8m}{c}\,(1 - e^{-(c/m)t})$$

Utilice Excel para resolver este problema con las variables de diseño r y n que minimicen el costo C.

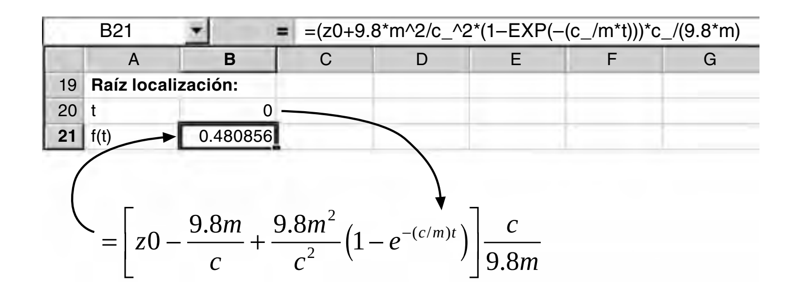
Solución Antes de llevar este problema a Excel, se debe enfrentar primero el problema de determinar la raíz en la formulación anterior [ecuación (E15.4.2)]. Un método podría ser desarrollar un macro para implementar un método de localización de raíces, tal como el de la bisección o de la secante. (Se ilustrará cómo realizar esto en el próximo capítulo, en la sección 16.3.)

Aunque, hay un procedimiento más fácil mediante la siguiente solución de la ecuación (E15.4.2) que es la iteración de punto fijo,

$$t_{i+1} = \left[500 + \frac{9.8m^2}{c^2}(1 - e^{-(c/m)t_i})\right]\frac{c}{9.8m} \tag{E15.4.3}$$

Así, t se ajusta hasta que se satisfaga la ecuación (E15.4.3). Se puede mostrar que para el rango de los parámetros usados en este problema, la fórmula siempre converge.

Ahora, ¿cómo se puede resolver esta ecuación en una hoja de cálculo? Como se muestra abajo, se fijan dos celdas para que tengan un valor de t y el lado derecho de la ecuación (E15.4.3) [es decir, $f(t)$].



Se puede teclear la ecuación (E15.4.3) en la celda B21 de tal forma que tome el valor del tiempo en la celda B20 y los valores de los otros parámetros se asignan en otras celdas en cualquier otro lugar de la hoja (véase a continuación cómo se construye toda la hoja). Después colóquese en la celda B20 y lleve su valor a la celda B21.

	A	B
19	**Raíz localización:**	
20	t	10.2551
21	f(t)	10.25595

Una vez que se introduce esta fórmula, se desplegará en forma inmediata el mensaje de error: "No se pueden resolver referencias circulares", ya que B20 depende de B21 y viceversa. Ahora, escoja del menú Herramientas/Opciones y seleccione **calculation** (cálculos). Del cuadro de diálogo "cálculos", escoja "iteración" y presione "aceptar". En forma inmediata la hoja de cálculo iterará estas celdas y el resultado aparecerá como la figura de la izquierda.

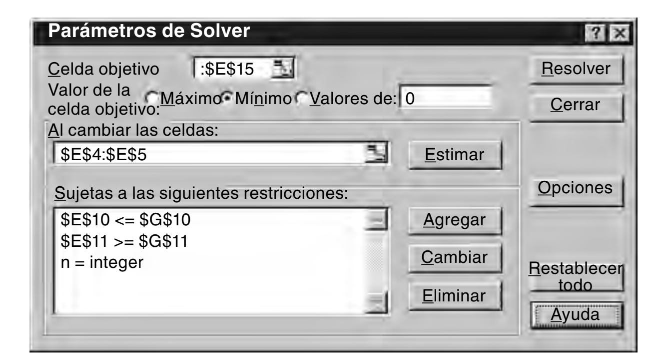
Así, las celdas convergirán a la raíz. Si se quiere tener más precisión, sólo presione la tecla F9 para que se realicen más iteraciones (por default son 100 iteraciones, que es posible modificar si así se desea).

En la figura 15.6 se muestra cómo implementar una hoja de cálculo en Excel para calcular los valores pertinentes. Las celdas no sombreadas son las que contienen los datos numéricos y las leyendas. Las celdas sombreadas contienen cantidades que se calculan basadas en las otras celdas. Por ejemplo, la masa en B17 se calculó con la ecuación (E15.4.1) con base en los valores de M_t (B4) y n (E5). Observe también que algunas celdas son redundantes. Por ejemplo, la celda E11 se refiere a la celda E5. Esta repetición en la celda E11 es para que la estructura de las restricciones sea evidente en la hoja. Finalmente, note que la celda que habrá de minimizarse es E15, que contiene el costo total. Las celdas que cambian son E4:E5, en las cuales se tiene el radio y el número de paracaídas.

	A	B	C	D	E	F	G
1	**Problema de optimización del paracaídas**						
2							
3	**Parámetros:**			**Variables de diseño:**			
4	Mt	2000		r	1		
5	g	9.8		n	1		
6	cost1	200					
7	cost2	56		**Restricciones:**			
8	cost3	0.1					
9	vc	20		**variable**		**tipo**	**límite**
10	kc	3		v	**95.8786**	<=	20
11	z0	500		n	**1**	>=	1
12							
13	**Valores calculados:**			**Función objetivo:**			
14	A	**6.283185**					
15	l	**1.414214**		Costo	**283.1438**		
16	c	**18.84956**					
17	m	**2000**					
18							
19	**Raíz localización:**						
20	t	**10.26439**					
21	f(t)	**10.26439**					

Figura 15.6
Hoja de cálculo en Excel que muestra la solución del problema de optimización no lineal del paracaídas.

Una vez que se ha creado la hoja de cálculo, se elige la solución Solver del menú herramientas. En esta etapa se desplegará un cuadro de diálogo, solicitándole la información pertinente. Las celdas en el cuadro de diálogo de Solver se llenan como se indica a la derecha.



Se deben agregar las restricciones una por una al seleccionar el botón "agregar". Esto abrirá un cuadro de diálogo que se ve así:



Como se muestra, la restricción de que la velocidad de impacto presente (celda E10) debe ser menor o igual que la velocidad requerida (G10) puede agregarse como se muestra. Después de agregar cada restricción se puede seleccionar el botón "agregar". Observe que la flecha hacia abajo le permite elegir entre varios tipos de restricciones (<=, >=, = y entero). Así, es posible forzar el número del paracaídas (E5) para que sea un entero.

Una vez introducidas las tres restricciones, se selecciona el botón "aceptar" para regresar al cuadro de diálogo de Solver. Después de seleccionar esta opción vuelva al menú de Solver. Cuando seleccione el botón "aceptar" se abrirá un cuadro de diálogo con un reporte sobre el éxito de la operación. En el caso presente, el Solver obtiene la solución correcta como se indica en la figura 15.7.

De esta forma, se determina que el costo mínimo es $4 377.26 si se divide la carga en seis paquetes con un radio del paracaídas de 2.944 m. Además de obtener la solución, el Solver también

	A	B	C	D	E	F	G
1	**Problema de optimización del paracaídas**						
2							
3	**Parámetros:**			**Variables de diseño:**			
4	Mt	2000		r	**2.943652**		
5	g	9.8		n	**6**		
6	cost1	200					
7	cost2	56		**Restricciones:**			
8	cost3	0.1					
9	vc	20		**variable**		**tipo**	**límite**
10	kc	3		v	**20**	<=	20
11	z0	500		n	**6**	>=	1
12							
13	**Valores calculados:**			**Función objetivo:**			
14	A	54.44435					
15	l	4.162953		Costo	**4377.264**		
16	c	163.333					
17	m	333.3333					
18							
19	**Raíz localización:**						
20	t	27.04077					
21	f(t)	27.04077					

Figura 15.7
Hoja de cálculo en Excel con la solución del problema de optimización no lineal referente al paracaídas.

proporciona algunos reportes en resumen útiles. Éstos se explorarán en las aplicaciones a la ingeniería que se describirán en la sección 16.2.

15.3.3 MATLAB

Como se resume en la tabla 15.1, MATLAB tiene varias funciones incorporadas para optimización. Los siguientes dos ejemplos ilustran cómo utilizarlas.

Tabla 15.1 Funciones de MATLAB para optimización.

Función	Descripción
`fminbnd`	Minimiza una función de una variable con restricciones
`fminsearch`	Minimiza una función de varias variables

EJEMPLO 15.5 Uso de MATLAB para la optimización unidimensional

Planteamiento del problema Utilice la función *`fminbnd`* de MATLAB para encontrar el máximo de

$$f(x) = 2 \text{ sen } x - \frac{x^2}{2}$$

en el intervalo $x_l = 0$ y $x_u = 4$. Recuerde que en el capítulo 13 empleamos varios métodos para resolver este problema para $x = 1.7757$ y $f(x) = 1.4276$.

Solución Primero, necesitamos crear un archivo fx.m para la función.

```
function f=fx(x)
f = -(2*sin(x)-x^2/10)
```

Como lo que nos interesa es la maximización, introducimos el negativo de la función. A continuación llamamos a la función `fminbnd` con

```
[x,f]=fminbnd('fx',0,4)
```

El resultado es

```
f =
   -1.7757
```

```
x =
   1.4275
```

Observe que se pueden incluir más argumentos. Una adición útil es establecer opciones de optimización, como tolerancia de error o máximo de iteraciones. Esto se hace con la función `optimset`, que se utilizó previamente en el ejemplo 7.6 y que tiene el formato general

```
optimset('param1',value1,'param2',value2,...)
```

donde $param_i$ es un parámetro que especifica el tipo de opción y $value_i$ es el valor asignado a esa opción. En el ejemplo si se quiere establecer la tolerancia de 1×10^{-2},

```
optimset('TolX',le–2)
```

De esta manera la solución del presente problema con una tolerancia de 1×10^{-2} se genera con

```
>> fminbnd('fx',0,4,optimset('TolX',le–2))
```

cuyo resultado es

```
f =
   –1.7757
ans =
     1.4270
```

Un conjunto completo de parámetros se encuentra llamando a “Help” (Ayuda)

```
>> Help optimset
```

MATLAB tiene diversas capacidades para manejar funciones multidimensionales. Recuerde del capítulo 13 que la imagen visual de una búsqueda unidimensional era la de una montaña rusa. Para casos bidimensionales, la imagen se vuelve la de montañas y valles. Como en el siguiente ejemplo, las capacidades gráficas de MATLAB proporcionan un medio cómodo para visualizar tales funciones.

EJEMPLO 15.6 Visualización de una función bidimensional

Planteamiento del problema Use las capacidades gráficas de MATLAB para mostrar la siguiente función y estimar visualmente su mínimo en el rango $-2 \le x_1 \le 0$ y $0 \le x_2 \le 3$:

$$f(x_1, x_2) = 2 + x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$$

Solución El siguiente código genera gráficas de contorno y malla de la función:

```
x=linspace(-2,0,40);y=linspace(0,3,40);
[X,Y] = meshgrid(x,y);
Z=2+X-Y+2*X.^2+2*X.*Y+Y.^2;
subplot(1,2,1);
cs=contour(X,Y,Z);clabel(cs);
xlabel('x_1');ylabel('x_2');
title('(a) Contour plot');grid;
subplot(1,2,2);
cs=surfc(X,Y,Z);
zmin=floor(min(Z));
zmax=ceil(max(Z));
xlabel('x_1');ylabel('x_2');zlabel('f(x_1,x_2)');
title('(b) Mesh plot');
```

Como se muestra en la figura 15.8, ambas gráficas indican que la función tiene un valor mínimo de alrededor de $f(x_1, x_2) = 0$ a 1, ubicado alrededor de $x_1 = -1$ y $x_2 = 1.5$.

El MATLAB estándar tiene una función `fminsearch` que se puede usar para determinar el mínimo de una función multidimensional. Se basa en el método Nelder-Mead, un método de búsqueda directa que sólo usa valores de función (no necesita derivadas) y maneja funciones objetivo no continuas. Una expresión sencilla de su sintaxis es

```
[xmin, fval] = fminsearch(function,x1,x2)
```

Donde `xmin` y `fval` son la ubicación y el valor del mínimo, `function` es el nombre de la función que se está evaluando, y `x1` y `x2` son los límites del intervalo que se está explorando.

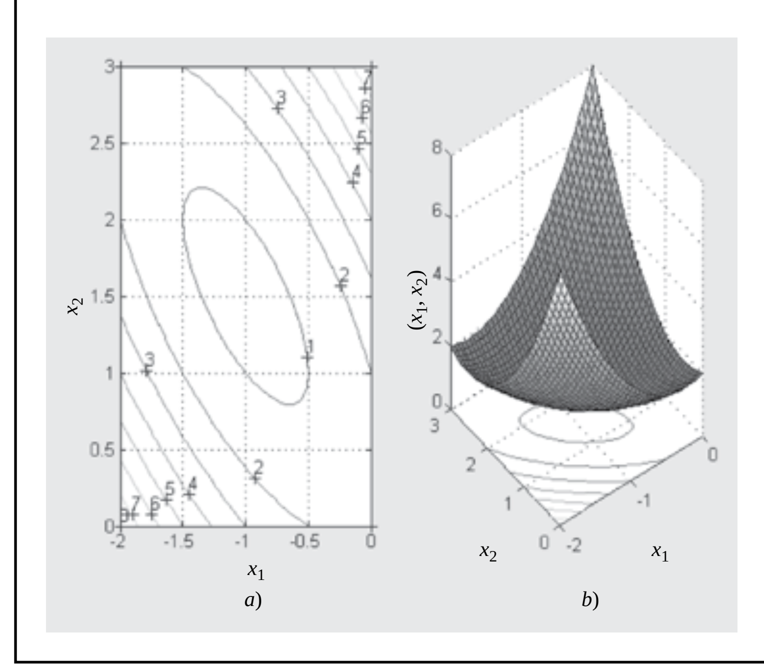


Figura 15.8
Gráficas *a*) de contorno y *b*) de malla de una función bidimensional.

EJEMPLO 15.7 Uso de MATLAB para optimización multidimensional

Planteamiento del problema Use la función `fminsearch` de MATLAB para encontrar el máximo para la función simple que se graficó en el ejemplo 15.6.

$$f(x_1, x_2) = 2 + x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$$

Utilice valores iniciales de $x = -0.5$ y $y = 0.5$.

Solución Se puede llamar la función `fminsearch` con

```
>> f=@(x) 2+x(1)-x(2)+2*x(1)^2+2*x(1)*x(2)+x(2)^2;
>> [x,fval]=fminsearch(f,[-0.5,0.5])

x =
   -1.0000    1.5000
fval =
    0.7500
```

Igual que con `fminbnd`, se pueden incluir argumentos con objeto de especificar parámetros adicionales del proceso de optimización. Por ejemplo, la función `optimset` se puede usar para limitar el número máximo de iteraciones

```
>> [x,fval]=fminsearch(f,[-0.5,0.5],optimset('MaxIter',2))
```

con el resultado

```
Exiting: Maximum number of iterations has been exceeded
         - increase MaxIter option.
         Current function value: 1.225625
x =
   -0.5000    0.5250
fval =
    1.2256
```

Así, como se ha establecido un límite muy estricto a las iteraciones, la optimización termina bastante antes de llegar al máximo.

15.3.4 Mathcad

Mathcad contiene una función de modo numérico que se llama **Find**, la cual puede usarse para resolver hasta 50 ecuaciones algebraicas simultáneas no lineales con restricciones de desigualdad. El uso de esta función para aplicaciones sin restricciones se describió en la parte dos. Si **Find** no logra ubicar una solución que satisfaga las ecuaciones y las restricciones, devuelve al usuario el mensaje de error “Did not find solution”. Sin embargo, Mathcad también contiene una función similar que se llama **Minerr**. Ésta da resultados de solución que minimizan los errores en las restricciones, incluso cuando no se pueden encontrar soluciones exactas. Esta función resuelve ecuaciones y acepta diversas restricciones usando el método Levenberg-Marquardt, tomado de los algoritmos de dominio público MINPACK, desarrollados y publicados por el Argonne National Laboratory.

Ahora se va a desarrollar un ejemplo en el que se usa **Find** para resolver un sistema de ecuaciones no lineales con restricciones. Se introducen valores iniciales de $x = -1$ y $y = 1$ usando el símbolo de definición como se muestra en la figura 15.9. La palabra **Given** alerta entonces a Mathcad de que lo que viene en seguida es un sistema de ecuaciones. Entonces se pueden introducir las ecuaciones y la restricción de desigualdad. Observe que, para esta aplicación, Mathcad necesita usar un signo de igual simbólico (que se teclea como [Ctrl] =) y > para separar los lados izquierdo y derecho de una ecuación. Ahora, el vector que consiste en `xval` y `yval` se calcula usando **Find**(x,y) y los valores se muestran empleando un signo de igual.

Se puede colocar en la hoja de cálculo una gráfica que muestre las ecuaciones y las restricciones, así como la solución, haciendo clic en la ubicación deseada. Esto pone una retícula roja en dicha

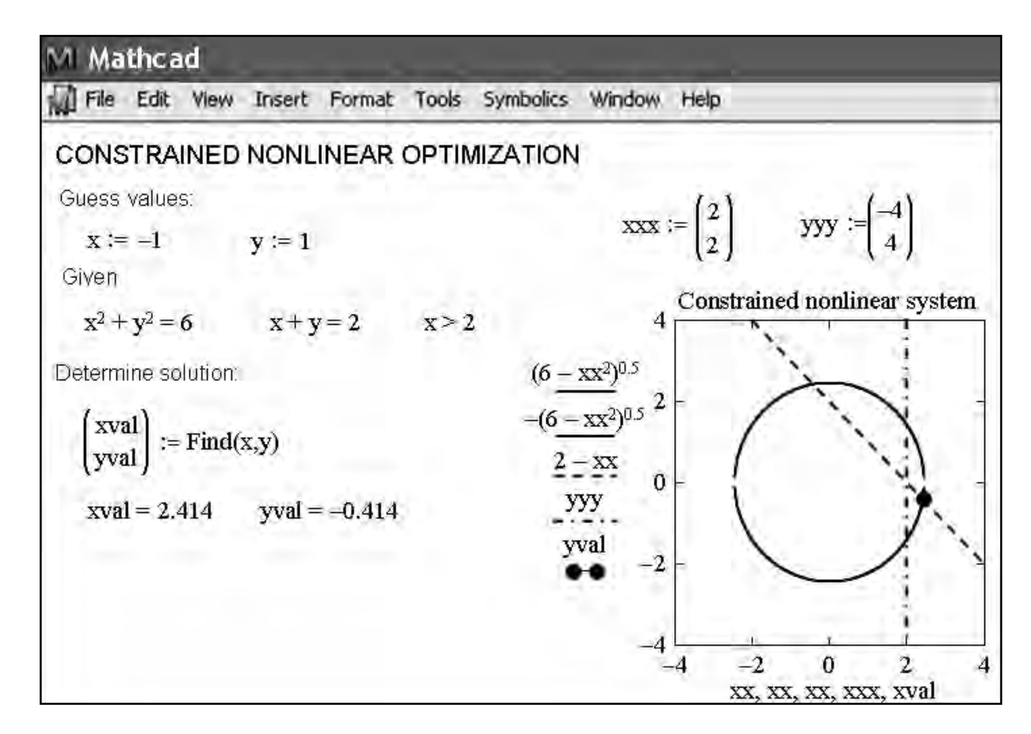


Figura 15.9
Pantalla Mathcad para un problema de optimización no lineal con restricciones.

ubicación. Luego use en la barra de menú Insert/Graph/X-Y Plot para colocar una gráfica en blanco en la hoja de cálculo con espacios para las expresiones que se graficarán y para los rangos de los ejes *x* y *y*. Se grafican cuatro variables en el eje *y* como se muestra: las mitades superior e inferior de la ecuación para la circunferencia, la función lineal y una línea vertical para representar la restricción $x > 2$. Además, se incluye la solución como un punto. Una vez que se ha creado la gráfica, se puede usar en la barra del menú Format/Graph/X-Y Plot para variar el tipo de gráfica, cambiar el color, el tipo y grosor de la traza de la función, y agregar títulos, rótulos y otras características. La gráfica y los valores numéricos para `xval` y `yval` ilustran muy bien la solución como la intersección de la circunferencia y la línea en la región en la que $x > 2$.

PROBLEMAS

15.1 Una compañía fabrica dos tipos de productos, A y B. Éstos se fabrican durante una semana laboral de 40 horas para enviarse al final de la semana. Se requieren 20 kg y 5 kg de materia prima por kilogramo de producto, respectivamente, y la compañía tiene acceso a 9 500 kg de materia prima por semana. Sólo se puede crear un producto a la vez, con tiempos de producción para cada uno de ellos de 0.04 y 0.12 horas, respectivamente. La planta sólo puede almacenar 550 kg en total de productos por semana. Por último, la compañía obtiene utilidades de \$45 y \$20 por cada unidad de A y B, respectivamente. Cada unidad de producto equivale a un kilogramo.

a) Plantee el problema de programación lineal para maximizar la utilidad.
b) Resuelva en forma gráfica el problema de programación lineal.
c) Solucione el problema de programación lineal con el método simplex.
d) Resuelva el problema con algún paquete de software.
e) Evalúe cuál de las opciones siguientes elevaría las utilidades al máximo: incrementar la materia prima, el almacenamiento, o el tiempo de producción.

15.2 Suponga que para el ejemplo 15.1, la planta de procesamiento de gas decide producir un tercer grado de producto con las características siguientes:

	Supremo
Gas crudo	15 m^3/ton
Tiempo de producción	12 h/ton
Almacenamiento	5 ton
Utilidad	\$250/ton

Además, suponga que se ha descubierto una nueva fuente de gas crudo, lo que duplicó el total disponible a 154 m^3/semana.

a) Plantee el problema de programación lineal para maximizar la utilidad.
b) Resuelva el problema de programación lineal con el método simplex.
c) Solucione el problema con un paquete de software.
d) Evalúe cuál de las opciones siguientes aumentaría las utilidades al máximo: incrementar la materia prima, el almacenamiento, o el tiempo de producción.

15.3 Considere el problema de programación lineal siguiente:

Maximizar $f(x, y) = 1.75x + 1.25y$

sujeta a:

$1.2x + 2.25y \leq 14$
$x + 1.1y \leq 8$
$2.5x + y \leq 9$
$x \geq 0$
$y \geq 0$

Obtenga la solución:

a) En forma gráfica.
b) Usando el método simplex.
c) Utilizando un paquete de software apropiado (por ejemplo, Excel, MATLAB o Mathcad).

15.4 Considere el problema de programación lineal que sigue:

Maximizar $f(x, y) = 6x + 8y$

sujeta a

$5x + 2y \leq 40$
$6x + 6y \leq 60$
$2x + 4y \leq 32$
$x \geq 0$
$y \geq 0$

Obtenga la solución:

a) En forma gráfica.
b) Usando el método simplex.
c) Utilizando un paquete de software apropiado (por ejemplo, Excel).

15.5 Emplee un paquete de software (por ejemplo, Excel, MATLAB o Mathcad) para resolver el problema siguiente de optimización no lineal con restricciones:

Maximizar $f(x, y) = 1.2x + 2y - y^3$

sujeta a

$2x + y \leq 2$
$x \geq 0$
$y \geq 0$

15.6 Utilice un paquete de software (por ejemplo, Excel, MATLAB o Mathcad) para resolver el siguiente problema de optimización no lineal con restricciones:

Maximizar $f(x, y) = 15x + 15y$

sujeta a

$x^2 + y^2 \leq 1$
$x + 2y \leq 2.1$
$x \geq 0$
$y \geq 0$

15.7 Considere el problema siguiente de optimización no lineal con restricciones:

Minimizar $f(x, y) = (x - 3)^2 + (y - 3)^2$

sujeta a

$$x + 2y = 4$$

a) Utilice el enfoque gráfico para estimar la solución.
b) Emplee un paquete de software (como Excel) para obtener una estimación más exacta.

15.8 Use un paquete de software para determinar el máximo de

$$f(x, y) = 2.25xy + 1.75y - 1.5x^2 - 2y^2$$

15.9 Emplee un paquete de software para determinar el máximo de

$$f(x, y) = 4x + 2y + x^2 - 2x^4 + 2xy - 3y^2$$

15.10 Dada la función siguiente,

$$f(x, y) = -8x + x^2 + 12y + 4y^2 + 2xy$$

use un paquete de software para determinar el mínimo:
a) En forma gráfica.
b) Numéricamente.
c) Sustituya el resultado del inciso *b*) en la función a fin de determinar el valor mínimo de $f(x, y)$.
d) Determine el Hessiano y su determinante, y sustituya el resultado del inciso *b*) para verificar que se detectó un mínimo.

15.11 Se le pide a usted que diseñe un silo cónico cubierto para almacenar 50 m^3 de desechos líquidos. Suponga que los costos de excavación son de \$100/m^3, los de cubrimiento lateral son de \$50/m^2 y los de la cubierta son de \$25/m^2. Determine las dimensiones del silo que minimizan el costo *a*) si la pendiente lateral no está restringida y *b*) la pendiente lateral debe ser menor de 45°.

15.12 Una compañía automotriz tiene dos versiones del mismo modelo de auto para vender, un cupé de dos puertas y otro de tamaño grande de cuatro puertas.
a) Encuentre gráficamente cuántos autos de cada diseño deben producirse a fin de maximizar la utilidad, y diga de cuánto es esta ganancia.
b) Con Excel, resuelva el mismo problema.

	Dos puertas	Cuatro puertas	Disponibilidad
Utilidad	\$13 500/auto	\$15 000/auto	
Tiempo de producción	15 h/auto	20 h/auto	8 000 h/año
Almacenamiento	400 autos	350 autos	
Demanda del consumidor	700/auto	500/auto	240 000 autos

15.13 Og es el líder de la tribu de cavernícolas *Calm Waters*, que está sorprendentemente avanzada en matemáticas, aunque con mucho atraso tecnológico. Él debe decidir acerca del número de mazos y hachas de piedra que deben producirse para la batalla próxima contra la tribu vecina de los *Peaceful Sunset*. La experiencia le ha enseñado que un mazo es bueno para generar en promedio 0.45 muertes y 0.65 heridas, en tanto que un hacha produce 0.70 muertes y 0.35 heridas. La producción de un mazo requiere 5.1 libras de piedra y 2.1 horas-hombre de trabajo, mientras que para un hacha se necesitan 3.2 libras de piedra y 4.3 horas-hombre de trabajo. La tribu de Og dispone de 240 libras de piedra para la producción de armas, y de un total de 200 horas-hombre de trabajo, antes de que pase el tiempo esperado para esta batalla (la cual, Og está seguro, pondrá fin para siempre a la guerra). Al cuantificar el daño que se inflige al enemigo, Og valora una muerte tanto como dos heridas, y desea producir la mezcla de armas que maximice el daño.

a) Formule la situación como un problema de programación lineal. Asegúrese de definir las variables de decisión.
b) Represente este problema en forma gráfica, y asegúrese de identificar todos los puntos de esquina factibles, así como los no factibles.
c) Resuelva el problema de forma gráfica.
d) Solucione el problema con el uso de una computadora.

15.14 Desarrolle un archivo M que esté expresamente diseñado para ubicar un máximo con el algoritmo de búsqueda de sección dorada. En otras palabras, configure if de tal manera que encuentre directamente el máximo en vez de encontrar el mínimo de $-f(x)$. Pruebe su programa con el mismo problema que el del ejemplo 13.1. La función debe tener las siguientes características:
- Iterar hasta que el error relativo caiga por debajo de un criterio de detención o se exceda un número máximo de iteraciones.
- Devolver tanto la *x* óptima como $f(x)$.

15.15 Desarrolle un archivo M para ubicar un mínimo con la búsqueda de sección dorada. En vez de usar los criterios estándar de detención (como en la figura 13.5), determine el número de iteraciones necesarias para alcanzar una tolerancia deseada.

15.16 Desarrolle un archivo M para implementar la interpolación parabólica y ubicar un mínimo. Pruebe su programa con el mismo problema que en el ejemplo 13.2. La función debe tener las siguientes características:
- Estar basada en dos valores iniciales y haga que el programa genere el tercer valor inicial en el punto medio del intervalo.
- Verificar si los valores abarcan un máximo. Si no, la función no debe implementar el algoritmo, sino devolver un mensaje de error.
- Iterar hasta que el error relativo caiga por debajo del criterio de detención o se exceda un número máximo de iteraciones.
- Devolver tanto *x* óptima como $f(x)$.
- Usar un enfoque cerrado (como en el ejemplo 13.2) para reemplazar valores viejos por nuevos.

15.17 La longitud de la escalera que puede resolver la esquina que se describe en la figura P15.17 se puede determinar calculando el valor de θ que minimiza la siguiente función:

$$L(\theta) = \frac{w_1}{\operatorname{sen}\theta} + \frac{w_2}{\operatorname{sen}(\pi - \alpha - \theta)}$$

Para el caso en el que $w_1 = w_2 = 2$ m, utilice un método numérico (incluyendo software) para desarrollar una gráfica de *L* contra una gama de α de 45° a 135°.

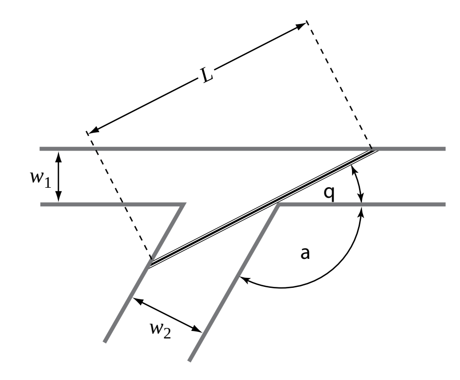


Figura P15.17
Una escalera resolviendo el paso por una esquina formada por dos pasillos.

CAPÍTULO 16

Estudio de casos: optimización

El propósito de este capítulo es utilizar los métodos numéricos analizados en los capítulos 13 al 15 para resolver problemas prácticos de ingeniería que involucren optimización. Estos problemas son importantes, ya que a los ingenieros con frecuencia se les pide que den la "mejor" solución a un problema. Como muchos de estos casos implican sistemas complicados e interacciones, entonces los métodos numéricos y las computadoras son necesarios para desarrollar soluciones óptimas.

Las siguientes aplicaciones son típicas de aquellas que se encuentran en forma rutinaria durante los estudios superiores y de graduados. Además, son representativas de problemas con los que se enfrentará el ingeniero profesionalmente. Los problemas se toman de las áreas de la ingeniería siguientes: química/bioingeniería, civil/ambiental, eléctrica y mecánica/aeronáutica.

La primera aplicación, tomada de la *ingeniería química/bioingeniería*, tiene que ver con el uso de la optimización no lineal con restricciones para el diseño óptimo de un tanque cilíndrico. Se usa el Solver de Excel para encontrar la solución.

Después, se utiliza la programación lineal para resolver un problema de la *ingeniería civil/ambiental*: minimizar el costo del tratamiento de aguas residuales para cumplir con los objetivos de calidad del agua en un río. En este ejemplo, se expone la noción de los precios indefinidos y su uso para mostrar la sensibilidad de una solución en programación lineal.

La tercera aplicación, tomada de la *ingeniería eléctrica*, implica maximizar la potencia a través de un potenciómetro en un circuito eléctrico. La solución involucra optimización sin restricciones unidimensional. Además de resolver el problema, se muestra cómo el lenguaje macro de Visual Basic permite el acceso al algoritmo de búsqueda de la sección dorada, dentro del contexto del ambiente Excel.

Finalmente, la cuarta aplicación, tomada de la *ingeniería mecánica/aeroespacial*, incluye la determinación de la posición de equilibrio de un sistema de resortes múltiples con base en la energía potencial mínima.

16.1 DISEÑO DE UN TANQUE CON EL MENOR COSTO (INGENIERÍA QUÍMICA/BIOINGENIERÍA)

Antecedentes Los ingenieros químicos (así como otros especialistas como los ingenieros mecánicos y civiles) con frecuencia se enfrentan al problema general del diseño de recipientes que transporten líquidos o gases. Suponga que se le pide determinar las dimensiones de un tanque cilíndrico pequeño para el transporte de desechos tóxicos que se van a trasladar en un camión. Su objetivo general será minimizar el costo del tanque. Sin embargo, además del costo, usted debe asegurar que pueda contener la cantidad requerida de líquido y que no exceda las dimensiones de la caja del camión. Debido a que el tanque transportará desechos tóxicos, se requiere que éste sea de un espesor determinado, dentro de ciertas regulaciones.

Un esquema del tanque y de la caja se muestra en la figura 16.1. Como se observa, el tanque es un cilindro con dos placas soldadas en cada extremo.

El costo del tanque tiene dos componentes: 1) gastos del material, que están basados en el peso, y 2) gastos de soldadura que se basan en la longitud a soldar. Note que esto último consiste en soldar tanto la junta interior como la junta exterior donde las placas se unen con el cilindro. Los datos necesarios para el problema se resumen en la tabla 16.1.

Tabla 16.1 Parámetros para determinar las dimensiones óptimas de un tanque cilíndrico para transporte de desechos tóxicos.

Parámetro	Símbolo	Valor	Unidades
Volumen requerido	V_o	0.8	m^3
Espesor	t	3	cm
Densidad	ρ	8 000	kg/m^3
Longitud de la caja	$L_{máx}$	2	m
Ancho de la caja	$D_{máx}$	1	m
Costo del material	c_m	4.5	\$/kg
Costo de soldadura	c_w	20	\$/m

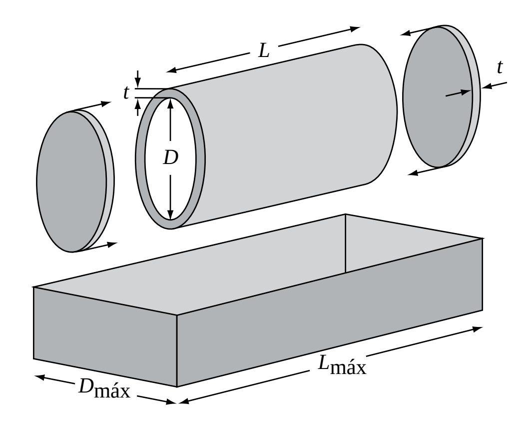


Figura 16.1
Parámetros para determinar las dimensiones óptimas de un tanque cilíndrico.

Solución El objetivo aquí es construir un tanque a un costo mínimo. El costo está relacionado con las variables de diseño (longitud y diámetro), ya que tienen efecto sobre la masa del tanque y las longitudes a soldar. Además, el problema tiene restricciones, pues el tanque debe 1) caber en la caja del camión y 2) tener capacidad para el volumen requerido de material.

El costo se obtiene de los costos del material del tanque y de la soldadura. Por lo tanto, la función objetivo se formula como una minimización

$$C = c_m m + c_w \ell_w \tag{16.1}$$

donde C = costo (\$), m = masa (kg), ℓ_w = longitud a soldar (m), c_m y c_w = factores de costo por masa (\$/kg) y longitud de soldadura (\$/m), respectivamente.

Después, se relacionan la masa y la longitud de soldadura con las dimensiones del tambor. Primero, se calcula la masa como el volumen del material por su densidad. El volumen del material usado para construir las paredes laterales (es decir, el cilindro) se calcula así:

$$V_{\text{cilindro}} = L\pi\left[\left(\frac{D}{2}+t\right)^2 - \left(\frac{D}{2}\right)^2\right]$$

Para cada placa circular en los extremos,

$$V_{\text{placa}} = \pi\left(\frac{D}{2}+t\right)^2 t$$

Así, la masa se calcula mediante

$$m = \rho\left\{L\pi\left[\left(\frac{D}{2}+t\right)^2 - \left(\frac{D}{2}\right)^2\right] + 2\pi\left(\frac{D}{2}+t\right)^2 t\right\} \tag{16.2}$$

donde ρ = densidad (kg/m^3).

La longitud de soldadura para unir cada placa es igual a la circunferencia interior y exterior del cilindro. Para las dos placas, la longitud total de soldadura será

$$\ell_w = 2\left[2\pi\left(\frac{D}{2}+t\right) + 2\pi\frac{D}{2}\right] = 4\pi(D+t) \tag{16.3}$$

Dados los valores para D y L (recuerde que el espesor t es fijado por un reglamento), las ecuaciones (16.1), (16.2) y (16.3) ofrecen un medio para calcular el costo. También observe que cuando las ecuaciones (16.2) y (16.3) se sustituyen en la ecuación (16.1), la función objetivo que se obtiene es no lineal.

Después, se formulan las restricciones. Primero, se debe calcular el volumen que el tanque terminado puede contener,

$$V = \frac{\pi D^2}{4} L$$

Este valor debe ser igual al volumen deseado. Así, una restricción es

$$\frac{\pi D^2 L}{4} = V_o$$

donde V_o es el volumen deseado (m^3).

Las restricciones restantes tienen que ver con que el tanque se ajuste a las dimensiones de la caja del camión,

$$L \le L_{\text{máx}}$$
$$D \le D_{\text{máx}}$$

El problema ahora está especificado. Con la sustitución de los valores de la tabla 16.1, se resume como

Maximizar $C = 4.5m + 20\ell_w$

sujeto a

$$\frac{\pi D^2 L}{4} = 0.8$$
$$L \le 2$$
$$D \le 1$$

donde

$$m = 8\,000\left\{L\pi\left[\left(\frac{D}{2}+0.03\right)^2 - \left(\frac{D}{2}\right)^2\right] + 2\pi\left(\frac{D}{2}+0.03\right)^2 0.03\right\}$$

y

$$\ell_w = 4\pi\,(D + 0.03)$$

El problema ahora se puede resolver de diferentes formas. Sin embargo, el método más simple para un problema de esta magnitud consiste en utilizar una herramienta como el Solver de Excel. La hoja de cálculo para realizar esto se muestra en la figura 16.2.

	A	B	C	D	E	F	G
1	**Diseño del tanque óptimo**						
2							
3	**Parámetros:**			**Variables de diseño:**			
4	V0	0.8		D	1		
5	t	0.03		L	2		
6	rho	8000					
7	Lmáx	2		**Restricciones:**			
8	Dmáx	1		D	**1**	<=	1
9	cm	4.5		L	**2**	<=	2
10	cw	20		Vol	**1.570796**	=	0.8
11							
12	**Valores calculados:**			**Función objetivo:**			
13	m	**1976.791**		C	**9154.425**		
14	lw	**12.94336**					
15							
16	Vcoraza	**0.19415**					
17	Vtapas	**0.052948**					

Figura 16.2
Hoja de cálculo de Excel lista para evaluar el costo de un tanque sujeto a restricciones de volumen y tamaño.

En el caso mostrado, se introducen los límites superiores para D y L. En este caso, el volumen es mayor que el requerido (1.57 > 0.8).

Una vez creada la hoja de cálculo, la selección Solver se elige del menú Tools (Herramientas). Aquí aparecerá una ventana de diálogo que le solicitará la información pertinente. Las celdas correspondientes para el cuadro de diálogo Solver se pueden llenar así



Al seleccionar el botón Resolver, un cuadro de diálogo se abrirá mostrando un reporte sobre el éxito de la operación. En el presente caso, Solver obtiene la solución correcta, la cual se muestra en la figura 16.3. Observe que el diámetro óptimo es casi el valor de la restricción de 1 m. Así, si aumentara la capacidad requerida del tanque, podría quitarse esta restricción y el problema se reduciría a una búsqueda unidimensional para la longitud.

	A	B	C	D	E	F	G
1	**Diseño del tanque óptimo**						
2							
3	**Parámetros:**			**Variables de diseño**			
4	V0	0.8		D	0.98351		
5	t	0.03		L	1.053033		
6	rho	8000					
7	Lmáx	2		**Restricciones**			
8	Dmáx	1		D	**0.98351**	<=	1
9	cm	4.5		L	**1.053033**	<=	2
10	cw	20		Vol	**0.799999**	=	0.8
11							
12	**Valores calculados:**			**Función objetivo:**			
13	m	**1215.206**		C	**5723.149**		
14	lw	**12.73614**					
15							
16	Vcoraza	**0.100587**					
17	Vtapas	**0.051314**					

Figura 16.3
Resultados de la minimización. El precio se reduce de \$9 154 a \$5 723, debido al menor volumen con dimensiones $D = 0.98$ m y $L = 1.05$ m.

16.2 MÍNIMO COSTO PARA EL TRATAMIENTO DE AGUAS RESIDUALES (INGENIERÍA CIVIL/AMBIENTAL)

Antecedentes Las descargas de aguas residuales de las grandes ciudades son, con frecuencia, la causa principal de la contaminación en un río. La figura 16.4 presenta el tipo de sistema que un ingeniero ambiental podría enfrentar. Varias ciudades están localizadas en las orillas de un río y sus afluentes. Cada una genera contaminación a una razón de carga P en unidades de miligramos por día (mg/d). La carga contaminante está sujeta al tratamiento de desechos que resultan de una remoción fraccional x. Así, la cantidad descargada al río es el exceso no removido por el tratamiento,

$$W_i = (1 - x_i)P_i \tag{16.4}$$

donde W_i = descarga de desechos de la i-ésima ciudad.

Cuando las descargas de desechos entran en la corriente, se mezclan con los contaminantes de las fuentes corriente arriba. Si se supone un mezclado completo en el punto de descarga, la concentración resultante en el punto de descarga se calcula con un simple balance de masa,

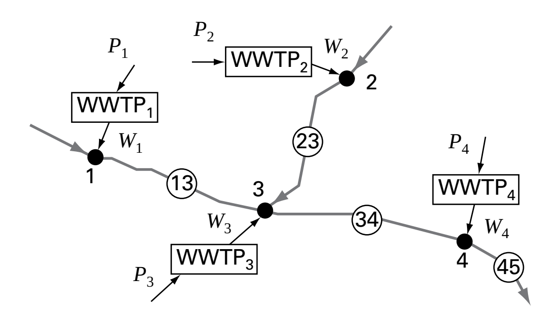


Figura 16.4
Cuatro plantas de tratamiento de aguas residuales que descargan contaminantes a un sistema de ríos. Los segmentos del río entre las ciudades están marcados con números dentro de un círculo.

$$c_i = \frac{W_i + Q_u c_u}{Q_i} \tag{16.5}$$

donde Q_u = flujo (L/d), c_u = concentración (mg/L) en el río corriente arriba de la descarga y Q_i = flujo corriente abajo del punto de descarga (L/d).

Después de que se establece la concentración en el punto de mezclado, los procesos de descomposición químicos y biológicos pueden eliminar algo de contaminación, conforme fluye corriente abajo. En el presente caso, se supone que esta eliminación puede representarse por un simple factor de reducción R.

Suponiendo que las fuentes de agua (es decir, las ciudades 1 y 2 en el río mostrado antes) están libres de contaminantes, las concentraciones en los cuatro nodos se calculan así:

$$c_1 = \frac{(1-x_1)P_1}{Q_{13}} \tag{16.6}$$

$$c_2 = \frac{(1-x_2)P_2}{Q_{23}}$$

$$c_3 = \frac{R_{13}Q_{13}c_1 + R_{23}Q_{23}c_2 + (1-x_3)P_3}{Q_{34}}$$

$$c_4 = \frac{R_{34}Q_{34}c_3 + (1-x_4)P_4}{Q_{45}}$$

Después, se observa que el tratamiento de aguas tiene un costo diferente, d_i ($1 000/mg eliminado), en cada una de las instalaciones. Así, el costo total de tratamiento (sobre una base diaria) se calcula como

$$Z = d_1P_1x_1 + d_2P_2x_2 + d_3P_3x_3 + d_4P_4x_4 \tag{16.7}$$

donde Z es el costo total diario del tratamiento ($1 000/d).

La pieza final en la "decisión" son las regulaciones ambientales. Para proteger los usos benéficos del río (por ejemplo, paseos en bote, pesca, uso como balneario), las regulaciones indican que la concentración del río no debe exceder un estándar de calidad c_s en el agua.

En la tabla 16.2 se resumen los parámetros para el sistema de ríos de la figura 16.4. Observe que hay una diferencia en los costos de tratamiento entre las ciudades corriente arriba (1 y 2) y corriente abajo (3 y 4), debido a la naturaleza obsoleta de las plantas corriente abajo.

Tabla 16.2 Parámetros para las cuatro plantas de tratamiento de aguas residuales que descargan contaminantes a un sistema de ríos, junto con las concentraciones resultantes (c_i) para tratamiento cero. También se dan el flujo, el factor de remoción y los estándares para los segmentos del río.

Ciudad	P_i (mg/d)	d_i ($\$10^{-6}$/mg)	c_i (mg/L)	Segmento	Q (L/d)	R	c_s (mg/L)
1	1.00×10^9	2	100	1-3	1.00×10^7	0.5	20
2	2.00×10^9	2	40	2-3	5.00×10^7	0.35	20
3	4.00×10^9	4	47.3	3-4	1.10×10^8	0.6	20
4	2.50×10^9	4	22.5	4-5	2.50×10^8		20

La concentración se calcula con la ecuación (16.6) y el resultado se presenta en la columna sombreada, para el caso en que no se implementó tratamiento de residuos (es decir, donde todas las $x = 0$). Observe que el estándar de 20 mg/L se viola en todos los puntos de mezclado.

Utilice la programación lineal para determinar los niveles de tratamiento que satisfacen los estándares de calidad del agua a un costo mínimo. También evalúe el impacto al hacer el estándar más restringido debajo de la ciudad 3. Es decir, realice el mismo ejercicio; pero ahora con los estándares para los segmentos 3-4 y 4-5 disminuidos a 10 mg/L.

Solución Todos los factores antes mencionados se combinan en el siguiente problema de programación lineal:

$$\text{Minimizar } Z = d_1P_1x_1 + d_2P_2x_2 + d_3P_3x_3 + d_4P_4x_4 \tag{16.8}$$

sujeto a las siguientes restricciones

$$\frac{(1-x_1)P_1}{Q_{13}} \le c_{s1}$$

$$\frac{(1-x_2)P_2}{Q_{23}} \le c_{s2} \tag{16.9}$$

$$\frac{R_{13}Q_{13}c_1 + R_{23}Q_{23}c_2 + (1-x_3)P_3}{Q_{34}} \le c_{s3}$$

$$\frac{R_{34}Q_{34}c_3 + (1-x_4)P_4}{Q_{45}} \le c_{s4}$$

$$0 \le x_1, x_2, x_3, x_4 \le 1 \tag{16.10}$$

De esta forma, la función objetivo es para minimizar el costo del tratamiento [ecuación (16.8)] sujeto a la restricción de los estándares de calidad del agua que se deben satisfacer en todas las partes del sistema [ecuación (16.9)]. Además, el tratamiento no debe ser negativo o mayor que el 100% de remoción [ecuación (16.10)].

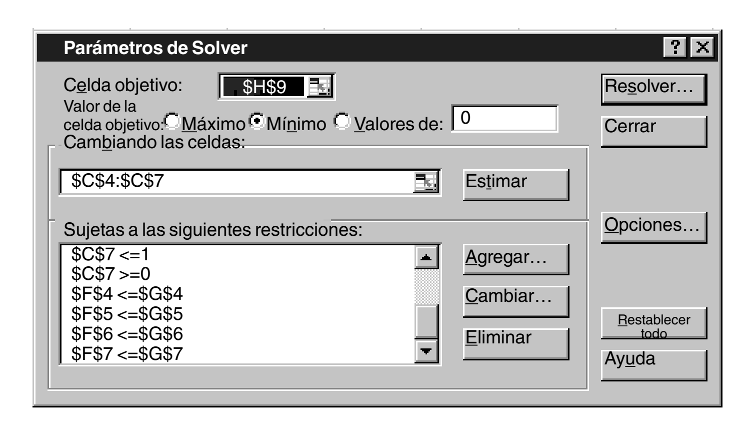
El problema se resuelve utilizando diversos paquetes. Para esta aplicación se utiliza la hoja de cálculo Excel. Como se observa en la figura 16.5, los datos junto con los cálculos de la concentración se pueden introducir fácilmente en las celdas de la hoja de cálculo.

	A	B	C	D	E	F	G	H
1	**Costo mínimo del tratamiento de aguas residuales**							
2		No tratada	Tratamiento	Descarga	Costo unit.	Concent.	Estándar	Costo de
3	Ciudad	P	x	W	d	en el río	de CA	tratamiento
4	1	1.00E+09	0	1.00E+09	2.00E-06	**100.00**	20.00	**0.00**
5	2	2.00E+09	0	2.00E+09	2.00E-06	40.00	20.00	0.00
6	3	4.00E+09	0	4.00E+09	4.00E-06	47.27	20.00	0.00
7	4	2.50E+09	0	2.50E+09	4.00E-06	22.48	20.00	0.00
8		Flujo en	Remoción					
9	Segmento	el río	en el río				Total	**0.00**
10	1-3	1.00E+07	0.5					
11	2-3	5.00E+07	0.35					
12	3-4	1.10E+08	0.6					
13	4-5	2.50E+08						

=D4/B10 **=$B$4*$C$4*$E$4** **=SUM(H4:H7)**

Figura 16.5
Hoja de cálculo de Excel lista para evaluar el costo del tratamiento de aguas en un sistema de ríos regulado. La columna F contiene el cálculo de la concentración de acuerdo con la ecuación (16.6). Las celdas F4 y H4 están resaltadas para mostrar las fórmulas usadas para calcular c_1 y el costo del tratamiento para la ciudad 1. Además, se resalta la celda H9 que muestra la fórmula para el costo total que es el que hay que minimizar [ecuación (16.8)].

Una vez que se crea la hoja de cálculo, se elige la selección Solver del menú Tools (Herramientas). En este punto, se desplegará una ventana de diálogo, solicitándole la información pertinente. Las celdas correspondientes para el cuadro de diálogo se podrían llenar así



Observe que no se muestran todas las restricciones, ya que el cuadro de diálogo despliega sólo seis restricciones a la vez.

Cuando se selecciona el botón Resolver, se abre un cuadro de diálogo con un reporte sobre el éxito de la operación. En el presente caso, Solver obtiene la solución correcta, la cual se muestra en la figura 16.6. Antes de aceptar la solución [al seleccionar el botón OK (aceptar) en el cuadro reporte del Solver], observe que se hayan generado 3 reportes: Respuesta, Sensibilidad y Límites. Seleccione el reporte Sensibilidad y después presione el botón OK para aceptar la solución. Solver generará automáticamente un reporte de sensibilidad, como el de la figura 16.7.

Ahora examinemos la solución (figura 16.6). Observe que el estándar será satisfecho en todos los puntos de mezclado. De hecho, la concentración en la ciudad 4 en realidad será menor que el estándar (16.28 mg/L), a pesar de que no se requerirá tratamiento para la ciudad 4.

	A	B	C	D	E	F	G	H
1	Costo mínimo del tratamiento de aguas residuales							
2		No tratada	Tratamiento	Descarga	Costo unit.	Concent.	Estándar	Costo del
3	Ciudad	P	x	W	d	en el río	de CA	tratamiento
4	1	1.00E+09	0.80	2.00E+08	2.00E-06	20.00	20.00	1600.00
5	2	2.00E+09	0.50	1.00E+09	2.00E-06	20.00	20.00	2000.00
6	3	4.00E+09	0.5625	1.75 E+09	4.00E-06	20.00	20.00	9000.00
7	4	2.50E+09	0.00	2.50E+09	4.00E-06	15.28	20.00	0.00
8		Flujo en	Remoción					
9	Segmento	el río	en el río				Total	12600.00
10	1-3	1.00E+07	0.5					
11	2-3	5.00E+07	0.35					
12	3-4	1.10E+08	0.6					
13	4-5	2.50E+08						

Figura 16.6
Resultados de la minimización. Los estándares de calidad del agua se satisfacen a un costo de $12 600/día. Observe que a pesar del hecho de que no se requiere tratamiento para la ciudad 4, la concentración en su punto de mezclado excede el estándar.

Como un ejercicio final, se pueden disminuir los estándares de 3-4 y 4-5 para tener 10 mg/L. Antes de hacerlo, se examina el reporte de sensibilidad. En el caso presente, la columna clave de la figura 16.7 es la de los multiplicadores de Lagrange (el precio anticipado). El *precio anticipado* es un valor que expresa la sensibilidad de la función objetivo (en nuestro caso, el costo) a una unidad de cambio de una de las restricciones (estándares de calidad-agua). Por lo tanto, representa el costo adicional en que se incurrirá al hacer más restrictivos los estándares. En nuestro ejemplo, es interesante que el precio anticipado mayor, $-\$440/\Delta c_{s3}$, se da para uno de los cambios de estándar (es decir, corriente abajo desde la ciudad 3) que se están contemplando. Lo anterior advierte que nuestra modificación será costosa.

Microsoft Excel 11.0 Reporte de sensibilidad
Hoja: [Sec1602.xls]Hoja1
Reporte elaborado: 3/28/2005 11:37:39 PM

Celdas ajustables

Celda	Nombre	Valor final	Disminución
$C$4	x	0.80	0.00
$C$5	x	0.50	0.00
$C$6	x	0.5625	0.00
$C$7	x	0.00	10000.00

Restricciones

Celda	Nombre	Valor final	Multiplicadores de Lagrange
$F$4	conc	20.00	0.00
$F$5	conc	20.00	-30.00
$F$6	conc	20.00	-440.00
$F$7	conc	15.28	0.00

Figura 16.7
Reporte de la sensibilidad en una hoja de cálculo para evaluar el costo del tratamiento de residuos en un sistema de ríos regulado.

Esto se confirma cuando se vuelve a ejecutar el Solver con los nuevos estándares (es decir, se disminuye el valor en las celdas G6 y G7 a 10). Como se muestra en la tabla 16.3, el resultado es que

el costo del tratamiento aumentó de $12 600/día a $19 640/día. Además, al reducir el estándar de concentraciones para las llegadas inferiores significará que la ciudad 4 debe comenzar a tratar sus desechos, y que la ciudad 3 debe actualizar su tratamiento. Note también que no se afecta el tratamiento en las ciudades corriente arriba.

Tabla 16.3 Comparación de dos escenarios que muestran el impacto de diferentes regulaciones sobre los costos de tratamiento.

Escenario 1: Todas las $c_s = 20$			Escenario 2: Corriente abajo $c_s = 10$		
Ciudad	***x***	***c***	**Ciudad**	***x***	***c***
1	0.8	20	1	0.8	20
2	0.5	20	2	0.5	20
3	0.5625	20	3	0.8375	10
4	0	15.28	4	0.264	10
Costo = $12 600			Costo = $19 640		

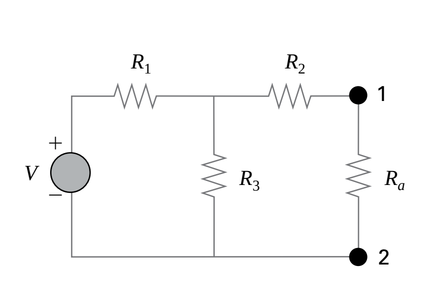


Figura 16.8
Un circuito de resistencias con un resistor ajustable o potenciómetro.

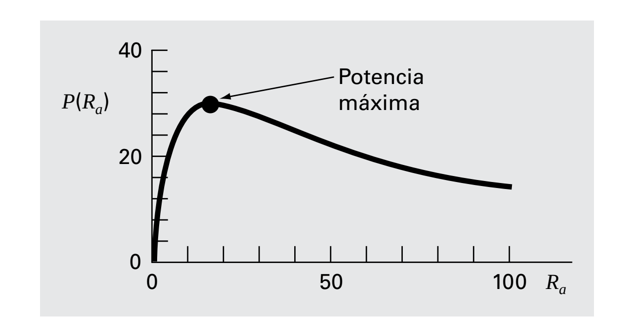


Figura 16.9
Una gráfica de transferencia de potencia a través de las terminales 1-2 de la figura 16.8, como una función de la resistencia del potenciómetro R_a.

16.3 MÁXIMA TRANSFERENCIA DE POTENCIA EN UN CIRCUITO (INGENIERÍA ELÉCTRICA)

Antecedentes El circuito de resistencias simple que se presenta en la figura 16.8 contiene tres resistores fijos y uno variable. Los resistores variables se llaman *potenciómetros*. Los valores de los parámetros son $V = 80$ V, $R_1 = 8\ \Omega$, $R_2 = 12\ \Omega$ y $R_3 = 10\ \Omega$. *a*) Encuentre el valor de la resistencia variable R_a que maximiza la transferencia de potencia a través de las terminales 1 y 2. *b*) Realice un análisis de sensibilidad para determinar cómo varían la máxima potencia y el valor correspondiente del potenciómetro (R_a) conforme V varía en un rango de 45 a 105 V.

Solución A partir de las leyes de Kirchhoff es posible obtener una expresión para la potencia del circuito:

$$P(R_a) = \frac{\left[\dfrac{VR_3R_a}{R_1(R_a + R_2 + R_3) + R_3R_a + R_3R_2}\right]^2}{R_a} \tag{16.11}$$

Sustituyendo los valores de los parámetros se obtiene la gráfica mostrada en la figura 16.9. Observe que la máxima transferencia de potencia se presenta con una resistencia de aproximadamente 16 Ω.

Resolveremos este problema de dos formas con la hoja de cálculo Excel. Primero, se emplean prueba y error y la opción Solver. Después, se desarrollará un programa macro en Visual BASIC, para realizar un análisis de sensibilidad.

a) En la figura 16.10 se muestra una hoja de cálculo de Excel para implementar la ecuación (16.11). Como se indica, la ecuación (16.11) se introduce en la celda B9. Entonces el valor de R_a (celda B8) se varía en forma de prueba y error hasta que se obtenga un residuo mínimo. En este ejemplo, el resultado es una potencia de 30.03 W con un valor en el potenciómetro de R_a = 16.44 Ω.

B9 = =(V*Res3*Ra/(Res1*(Ra+Res2+Res3)+Res3*Ra+Res3*Res2))^2/Ra

	A	B	C	D	E	F	G	H
1	**Máxima transferencia de potencia**							
2								
3	V	80						
4	Res1	8						
5	Res2	12						
6	Res3	10						
7								
8	Ra	16.444444						
9	P(Ra)	**30.03003**						
10								

Figura 16.10
Determinación en Excel de la máxima potencia a través de un potenciómetro mediante el uso de prueba y error.

Un procedimiento mejor consiste en utilizar la opción Solver del menú Tools (Herramientas) de la hoja de cálculo. Se desplegará una ventana de diálogo solicitándole la información pertinente. Las celdas correspondientes para el cuadro de diálogo Solver se llenarán así

Ubique la celda destino:	B9
Igual a ● máx ❍ mín ❍ igual a	0
Por cambio de celdas	B8

Cuando se selecciona el botón OK (aceptar), se despliega un cuadro de diálogo con un reporte sobre el éxito de la operación. En el caso actual, Solver obtiene la misma solución correcta que se presenta en la figura 16.10.

b) Ahora, aunque el procedimiento anterior es excelente para una sola evaluación, no es conveniente para los casos donde se emplean múltiples optimizaciones. Tal podría ser el caso en la segunda parte de esta aplicación, en la cual estamos interesados en determinar de qué modo la máxima potencia varía con diferentes valores de voltaje. En efecto, se podría llamar muchas veces el Solver con los diferentes valores de los parámetros; pero esto resultaría ineficiente. Sería preferible desarrollar una función macro que encuentre el óptimo.

Tal función se muestra en la figura 16.11. Advierta su similitud con el pseudocódigo de la búsqueda de la sección dorada que se presentó en la figura 13.5. Además, observe que una función se debe definir también para calcular la potencia de acuerdo con la ecuación (16.11).

En la figura 16.12 se muestra una hoja de cálculo Excel que utiliza este macro para evaluar la sensibilidad de la solución al voltaje. Se tiene una columna de valores que cubre los valores de los voltajes (es decir, de 45 a 105 V). En la celda B9 se tiene una función macro que referencia el valor adyacente de *V* (los 45 voltios en A9). Además, se dan también los otros parámetros en el argumento de la función. Advierta que, mientras la referencia a *V* es relativa, las referencias a los valores iniciales superior e inferior y a las resistencias son absolutas (es decir, incluyen el signo $). Esto se hizo de tal forma que cuando la fórmula se copie abajo, las referencias absolutas queden fijas; mientras que la referencia relativa corresponda al voltaje en el mismo renglón. Una estrategia similar se usa para introducir la ecuación (16.11) en la celda C9.

Cuando se copian las fórmulas hacia abajo, el resultado es como el que se presenta en la figura 16.12. La máxima potencia se puede graficar para visualizar el impacto de las variaciones de voltaje. En la figura 16.13 se observa que la potencia aumenta con el voltaje.

Los resultados de los valores correspondientes en el potenciómetro (R_a) son más interesantes. La hoja de cálculo indica que para un mismo valor, 16.44 Ω, da una máxima potencia. Tal resultado podría ser difícil de intuir basándose en una inspección de la ecuación (16.11).

```
Option Explicit

Function Golden(xlow, xhigh, R1, R2, R3, V)
Dim iter As Integer, maxit As Integer, ea As Double, es As Double
Dim fx As Double, xL As Double, xU As Double, d As Double, x1 as Double
Dim x2 As Double, f1 As Double, f2 As Double, xopt As Double
Const R As Double = (5 ^ 0.5 – 1) / 2
maxit = 50
es = 0.001
xL = xlow
xU = xhigh
iter = 1
d = R * (xU – xL)
x1 = xL + d
x2 = xU – d
f1 = f(x1, R1, R2, R3, V)
f2 = f(x2, R1, R2, R3, V)
If f1 > f2 Then
  xopt = x1
  fx = f1
Else
  xopt = x2
  fx = f2
End If
Do
  d = R * d
  If f1 > f2 Then
    xL = x2
    x2 = x1
    x1 = xL + d
    f2 = f1
    f1 = f(x1, R1, R2, R3, V)
  Else
    xU = x1
    x1 = x2
    x2 = xU – d
    f1 = f2
    f2 = f(x2, R1, R2, R3, V)
  End If
  iter = iter + 1
  If f1 > f2 Then
    xopt = x1
    fx =f1
Else
    xopt = x2
    fx = f2
  End If
  If xopt <> 0 Then ea = (1 – R) * Abs((xU – xL) / xopt) * 100
  If ea <= es Or iter >= maxit Then Exit Do
Loop
Golden = xopt
End Function

Function f(Ra, R1, R2, R3, V)
f = (V * R3 * Ra / (R1 * (Ra + R2 + R3) + R3 * Ra + R3 * R2)) ^ 2 / Ra
End Function
```

Figura 16.11
Macro para Excel escrito en Visual BASIC que determina un máximo con la búsqueda de la sección dorada.

Figura 16.12
Hoja de cálculo de Excel para implementar un análisis de sensibilidad de la máxima potencia con variaciones de voltaje. Esta rutina accesa el programa macro para la búsqueda de la sección dorada de la figura 16.11.

	A	B	C	D
1	Máxima transferencia de potencia			
2				
3	R1	8		
4	R2	12		
5	R3	10		
6	Rmín	0.1		
7	Rmáx	100		
8	V	Ra	P(Ra)	
9	45	**16.44444**	**9.501689**	
10	60	16.44444	16.89189	
11	75	16.44444	26.39358	
12	90	16.44444	38.00676	
13	105	16.44444	51.73142	

16.4 EQUILIBRIO Y ENERGÍA POTENCIAL MÍNIMA (INGENIERÍA MECÁNICA/AEROESPACIAL)

Antecedentes Como en la figura 16.14*a*, se puede fijar a una pared un resorte sin carga. Cuando se aplica una fuerza horizontal, el resorte se estira. El desplazamiento está relacionado con la fuerza por la *ley de Hooke*: $F = kx$. La *energía potencial* del desplazamiento consiste en la diferencia entre la energía de tensión del resorte y el trabajo realizado por la fuerza,

$$PE(x) = 0.5kx^2 - Fx \tag{16.12}$$

La ecuación (16.12) define una parábola. Como la energía potencial estará en un mínimo en el equilibrio, la solución para hallar el desplazamiento se puede ver como un problema de optimización unidimensional. Como la ecuación es tan fácil de diferenciar, se puede despejar el desplazamiento con $x = F/k$. Por ejemplo, si $k = 2$ N/cm y $F = 5$ N, $x = 5$ N/(2 N/cm)/5 = 2.5 cm.

En la figura 16.15 se muestra un caso bidimensional más interesante. En este sistema, hay dos grados de libertad porque el sistema se puede mover tanto horizontal como verticalmente. Del mismo modo que en el que se trató el sistema unidimensional, las deformaciones en el equilibrio son los valores de x_1 y x_2 que minimizan la energía potencial,

$$PE(x_1, x_2) = 0.5k_a\left(\sqrt{x_1^2 + (L_a - x_2)^2} - L_a\right)^2 + 0.5k_b\left(\sqrt{x_1^2 + (L_b + x_2)^2} - L_b\right)^2 - F_1x_1 - F_2x_2 \tag{16.13}$$

Si los parámetros son $k_a = 9$ N/cm, $k_b = 2$ N/cm, $L_a = 10$ cm, $L_b = 10$ cm, $F_1 = 2$ N y $F_2 = 4$ N, despeje los desplazamientos y la energía potencial.

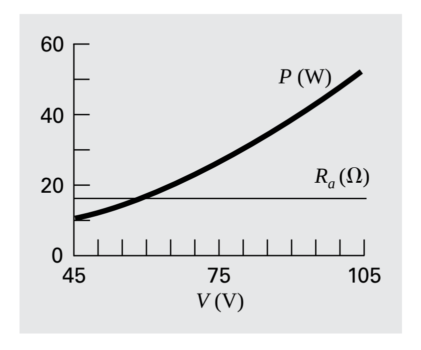


Figura 16.13
Resultados del análisis de sensibilidad del efecto de las variaciones de voltaje sobre la máxima potencia.

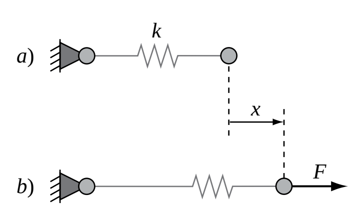


Figura 16.14
a) Un resorte sin carga fijado a una pared. *b*) La aplicación de una fuerza horizontal estira el resorte, donde la relación entre fuerza y desplazamiento se describe por la ley de Hooke.

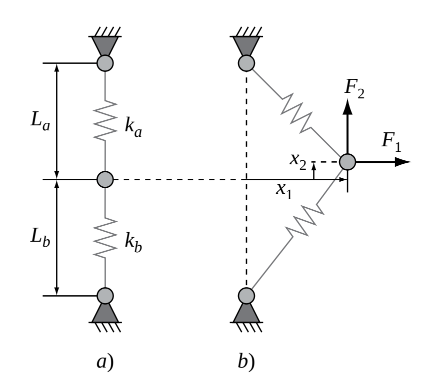


Figura 16.15
Un sistema de dos resortes: *a*) sin carga y *b*) con carga.

Antecedentes   Se pueden usar diversas herramientas de software para resolver este problema. Por ejemplo, usando MATLAB se puede desarrollar un archivo M que contenga la función de energía potencial,

```
function p=PE(x,ka,kb,La,Lb,F1,F2)
PEa=0.5*ka*(sqrt(x(1)^2+(La-x(2))^2)-La)^2;
PEb=0.5*kb*(sqrt(x(1)^2+(Lb+x(2))^2)-Lb)^2;
W=F1*x(1)+F2*x(2);
p=PEa+PEb-W;
```

La solución se puede entonces obtener con la función `fminsearch`

```
>> ka=9;kb=2;La=10;Lb=10;F1=2;F2=4;
>> [x,f]=fminsearch(@PE,[–0.5,0.5],[],ka,kb,La,Lb,F1,F2)
x =
   4.9523    1.2769
f =
   -9.6422
```

Así, en el equilibrio, la energía potencial es –9.6422 N · cm. El punto de conexión está ubicado 4.9523 cm a la derecha y 1.2759 cm arriba de su posición original.

PROBLEMAS

Ingeniería química/bioingeniería

16.1 Diseñe el contenedor cilíndrico óptimo (figura P16.1) de tal forma que abra por un extremo y tenga paredes de espesor despreciable. El contenedor va a almacenar 0.5 m^3. Realice el diseño de tal forma que el área del fondo y de sus lados sean mínimos.

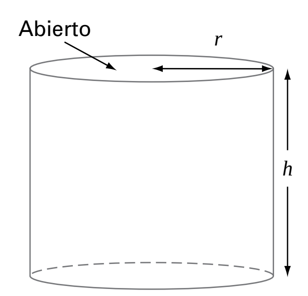


Figura P16.1
Un contenedor cilíndrico sin tapa.

16.2 *a*) Diseñe el contenedor cónico óptimo (figura P16.2) que tenga tapa y paredes de espesor despreciable. El contenedor debe tener un volumen de 0.5 m^3. Diséñelo de tal manera que las áreas de su tapa y lados se minimicen. *b*) Repita *a*), pero para un contenedor cónico sin tapa.

16.3 Diseñe el tanque cilíndrico óptimo con extremos convexos (figura P16.3). El contenedor debe tener capacidad de 0.5 m^3 y paredes de espesor despreciable. Observe que el área y el volumen de cada uno de los extremos convexos se puede calcular con

$$A = \pi(h^2 + r^2) \qquad V = \frac{\pi h(h^2 + 3r^2)}{6}$$

a) Diseñe el tanque de modo que su área superficial se minimice. Interprete el resultado.

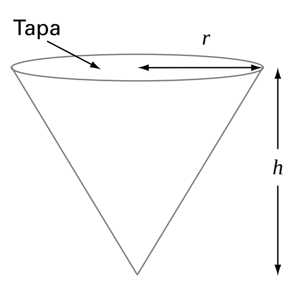


Figura P16.2
Un contenedor cónico con tapa.

b) Repita el inciso *a*), pero agregue la restricción $L \geq 2h$.

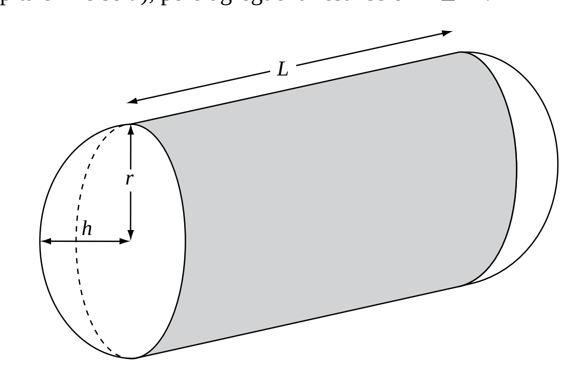


Figura P16.3

16.4 La razón de crecimiento específico de una levadura que produce un antibiótico es una función de la concentración del alimento *c*,

$$g = \frac{2c}{4 + 0.8c + c^2 + 0.2c^3}$$

Como se ilustra en la figura P16.4, el crecimiento parte de cero a muy bajas concentraciones debido a la limitación de la comida. También parte de cero en altas concentraciones debido a los efectos de toxicidad. Encuentre el valor de c para el cual el crecimiento es un máximo.

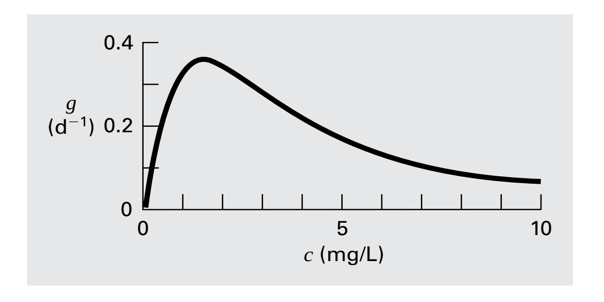


Figura P16.4
La razón de crecimiento específico de una levadura que produce un antibiótico contra la concentración de alimento.

16.5 Una planta química elabora sus tres productos principales en una semana. Cada uno de estos productos requiere cierta cantidad de materia prima química y de diferentes tiempos de producción, obteniéndose distintas utilidades. La información necesaria se resume en la tabla P16.5. Observe que hay suficiente espacio en la bodega de la planta para almacenar un total de 450 kg/semana.

a) Establezca un problema de programación lineal para maximizar las utilidades.
b) Resuelva el problema de programación lineal con el método simplex.
c) Resuelva el problema con un paquete de software.
d) Evalúe cuál de las siguientes opciones aumentará más las utilidades: incrementar la materia prima, el tiempo de producción o el almacenaje.

16.6 Recientemente los ingenieros químicos se han interesado en el área conocida como *minimización de desechos*. Ésta considera la operación de una planta química de modo tal que se minimicen los impactos sobre el ambiente. Suponga que una refinería desarrolla un producto, $Z1$, hecho de dos materias primas X y Y. La producción de 1 tonelada métrica del producto requiere 1 tonelada de X y 2.5 toneladas de Y y produce 1 tonelada de un líquido de desecho, W. Los ingenieros tienen tres alternativas para los desechos:

- Producir una tonelada de un producto secundario, $Z2$, al agregar una tonelada más de X por cada tonelada de W.
- Producir una tonelada de otro producto secundario, $Z3$, al agregar 1 tonelada más de Y por cada tonelada de W.
- Tratar los desechos de tal forma que su descarga sea permisible.

Los productos dan utilidades de $2 500, –$50 y $200/tonelada de $Z1$, $Z2$ y $Z3$, respectivamente. Observe que al producir $Z2$, de hecho, se obtiene una pérdida. El costo del proceso de tratamiento es de $300/tonelada. Además, la compañía tiene un límite de 7 500 y 10 000 toneladas de X y Y, respectivamente, durante el periodo de producción. Determine qué cantidad de productos y desechos se deben producir para maximizar las utilidades.

16.7 Hay que separar una mezcla de benceno y tolueno en un reactor flash. ¿A qué temperatura deberá operarse el reactor para obtener la mayor pureza de tolueno en la fase líquida (maximizar x_T)? La presión en el reactor es de 800 mm Hg. Las unidades en la ecuación de Antoine son mm Hg y °C para presión y temperatura, respectivamente.

$$x_B P_{\text{sat}_B} + x_T P_{\text{sat}_T} = P$$

$$\log_{10}(P_{\text{sat}_B}) = 6.905 - \frac{1\,211}{T + 221}$$

$$\log_{10}(P_{\text{sat}_T}) = 6.953 - \frac{1\,344}{T + 219}$$

16.8 A se convertirá en B en un reactor con agitación. El producto B y la sustancia sin reaccionar A se purifican en una unidad de separación. La sustancia A que no entró en la reacción se recicla al reactor. Un ingeniero de procesos ha encontrado que el costo inicial del sistema es una función de la conversión, x_A. Encuentre la conversión que dará el sistema de menor costo. C es una constante de proporcionalidad.

$$\text{Costo} = C\left[\left(\frac{1}{(1-x_A)^2}\right)^{0.6} + 6\left(\frac{1}{x_A}\right)^{0.6}\right]$$

16.9 En el problema 16.8 se utiliza sólo un reactor. Si se usan dos reactores en serie, cambia la ecuación que rige el sistema. Encuentre las conversiones en ambos reactores (x_{A1} y x_{A2}), de forma que se minimicen los costos totales del sistema.

Costo =

$$C\left[\left(\frac{x_{A1}}{x_{A2}(1-x_{A1})^2}\right)^{0.6} + \left(\frac{1-\left(\frac{x_{A1}}{x_{A2}}\right)}{(1-x_{A2})^2}\right)^{0.6} + 6\left(\frac{1}{x_{A2}}\right)^{0.6}\right]$$

16.10 En la reacción:

$$2A + B \Leftrightarrow C$$

el equilibrio se expresa como:

$$K = \frac{[C]}{[A]^2[B]} = \frac{[C]}{[A_0 - 2C]^2[B_0 - C]}$$

Si $K = 2\ M^{-1}$, se puede modificar la concentración inicial de A (A_0). La concentración inicial de B se fija por el proceso, $B_0 = 100$. A

Tabla P16.5

	Producto 1	Producto 2	Producto 3	Disponibilidad de fuentes
Materia prima química	6 kg/kg	4 kg/kg	13 kg/kg	3 000 kg
Tiempo de producción	0.05 h/kg	0.1 h/kg	0.2 h/kg	55 h/semana
Utilidad	$30/kg	$30/kg	$35/kg	

cuesta \$1/M y C se vende a \$10/M. ¿Cuál será la concentración inicial óptima de A que habrá de usarse de manera que se maximicen las utilidades?

16.11 Una planta química necesita 10^6 L/día de una solución. Se tienen tres fuentes con diferentes tasas de precios y suministros. Cada fuente tiene también concentraciones diferentes de una impureza que no debe rebasar cierto nivel, para evitar interferencias con las sustancias químicas. Los datos de las tres fuentes se resumen en la tabla siguiente. Determine la cantidad de cada fuente que satisfaga los requerimientos al menor costo.

	Fuente 1	Fuente 2	Fuente 3	Requerimiento
Costo (\$/L)	0.50	1.00	1.20	minimizar
Suministro (10^5 L/día)	20	10	5	≥10
Concentración (mg/L)	135	100	75	≤100

16.12 Usted tiene que diseñar un canal triangular abierto para transportar una corriente de desechos desde una planta química hasta un depósito de estabilización de desechos (figura P16.12). La velocidad media aumenta con el radio hidráulico, $R_h = A/p$, donde A es el área de la sección transversal y P es igual al perímetro mojado. Como la razón de flujo máximo corresponde a la velocidad máxima, el diseño óptimo tratará de minimizar el perímetro mojado. Determine las dimensiones que minimicen el perímetro mojado para un área dada de la sección transversal. ¿Las dimensiones están dentro de las medidas estándar?

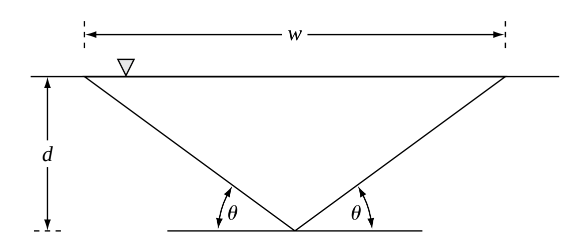


Figura P16.12

16.13 Un ingeniero agrícola tiene que diseñar un canal trapezoidal abierto para transportar el agua para irrigación (figura P16.13). Determine las dimensiones óptimas para minimizar el perímetro mojado en un área de sección transversal de 100 m^2. ¿Las dimensiones están dentro de las medidas estándar?

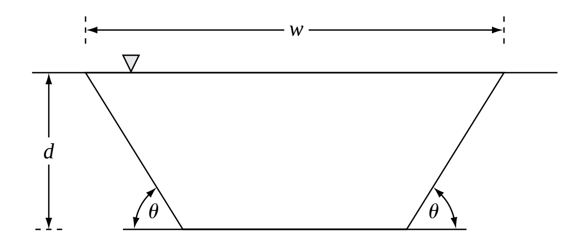


Figura P16.13

16.14 Calcule las dimensiones óptimas para un tanque cilíndrico térmico diseñado para contener 10 m^3 de fluido. Los extremos y laterales cuestan \$200/m^2 y \$100/m^2, respectivamente. Además, se aplica un recubrimiento a toda el área del tanque, la cual cuesta \$50/m^2.

Ingeniería civil/ambiental

16.15 Si se optimiza la ecuación siguiente se obtiene un modelo de elemento finito para una viga volada sujeta a cargas y momentos (figura P16.15)

$$f(x, y) = 5x^2 - 5xy + 2.5y^2 - x - 1.5y$$

donde x = desplazamiento final y y = momento final. Calcule los valores de x y y que minimizan $f(x, y)$.

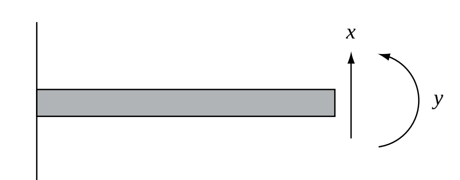


Figura P16.15
Viga volada.

16.16 Suponga usted que se le pide diseñar una columna que soporte una carga de compresión P, como se muestra en la figura P16.16*a*. La columna tiene una sección transversal en forma de tubo de pared delgada, como se aprecia en la figura P16.16*b*.

Las variables de diseño son el diámetro medio del tubo d y el espesor de la pared t. El costo del tubo se calcula por medio de la ecuación

$$\text{Costo} = f(t, d) = c_1W + c_2d$$

donde $c_1 = 4$ y $c_2 = 2$ son los factores de costo y W = peso del tubo,

$$W = \pi\ dt\ H\rho$$

donde ρ = densidad del material del tubo = 0.0025 kg/cm^3. La columna debe dar apoyo a la carga bajo un esfuerzo de compresión sin flexionarse. Por lo tanto,

$$\text{Esfuerzo real } (\sigma) \le \text{esfuerzo máximo de compresión}$$
$$= \sigma_y = 550 \text{ kg/cm}^2$$

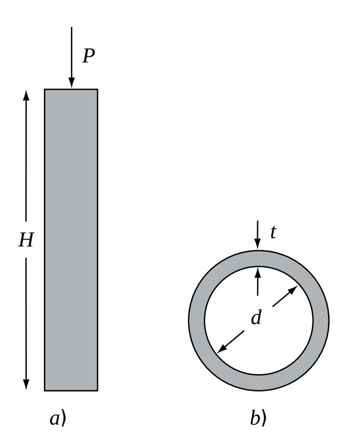


Figura P16.16
a) Una columna que soporta una carga de compresión *P*.
b) La columna tiene una sección transversal en forma de tubo de pared delgada.

Esfuerzo real ≤ esfuerzo de flexión

El esfuerzo real está dado por

$$\sigma = \frac{P}{A} = \frac{P}{\pi\ dt}$$

Se puede demostrar que el esfuerzo de flexión es

$$\sigma_b = \frac{\pi EI}{H^2 dt}$$

donde E = módulo de elasticidad e I = segundo momento del área de la sección transversal. Con cálculo se muestra que

$$I = \frac{\pi}{8} dt(d^2 + t^2)$$

Por último, los diámetros de los tubos disponibles se encuentran entre d_1 y d_2, y el espesor está entre t_1 y t_2. Desarrolle y resuelva este problema con la determinación de los valores de d y t que minimizan el costo. Obsérvese que H = 275 cm, P = 2 000 kg, E = 900 000 kg/cm^2, d_1 = 1 cm, d_2 = 10 cm, t_1 = 0.1 cm y t_2 = 1 cm.

16.17 El *modelo Streeter-Phelps* se utiliza para calcular la concentración de oxígeno disuelto en un río aguas abajo del punto de descarga de un drenaje (véase la figura P16.17),

$$o = o_s - \frac{k_d L_o}{k_d + k_s - k_a}(e^{-k_a t} - e^{-(k_d+k_s)t}) - \frac{S_b}{k_a}(1 - e^{-k_a t}) \qquad \text{(P.16.17)}$$

donde o = concentración del oxígeno disuelto [mg/L], o_s = concentración de saturación del oxígeno [mg/L], t = tiempo de travesía [d], L_o = concentración de la demanda bioquímica de oxígeno (DOB) en el punto de mezcla [mg/L], k_d = razón de descomposición de DOB [d^{-1}], k_s = razón de asentamiento de DBO [d^{-1}], k_a = razón de oxigenación [d^{-1}] y S_b = demanda de oxígeno sedimentario [mg/L/d].

Como se indica en la figura P16.17, la ecuación (P16.17) produce un "decaimiento" de oxígeno que alcanza un nivel mínimo crítico o_c para cierto tiempo de travesía t_c abajo del punto de descarga. Este punto se denomina "crítico" porque representa la ubicación en que la biota (flora y fauna) que depende del oxígeno (como los peces) estaría sujeta a la amenaza máxima. Determine el tiempo de travesía y la concentración críticos, dados los valores siguientes:

o_s = 10 mg/L	k_d = 0.1 d^{-1}	k_a = 0.6 d^{-1}
k_s = 0.05 d^{-1}	L_o = 50 mg/L	S_b = 1 mg/L/d

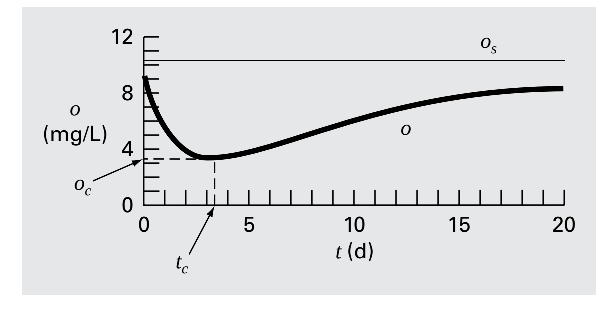


Figura P16.17
Un "decaimiento" de oxígeno disuelto debajo del punto de descarga de un drenaje hacia un río.

16.18 La distribución bidimensional de la concentración de cierto contaminante en un canal está descrita por la ecuación

$$c(x, y) = 7.9 + 0.13x + 0.21y - 0.05x^2 \\ -0.016y^2 - 0.007xy$$

Determine la ubicación exacta de la concentración máxima dada la función, si se sabe que se encuentra entre los límites de $-10 \leq x \leq 10$ y $0 \leq y \leq 20$.

16.19 El flujo Q [m^3/s] en un canal abierto se pronostica con la ecuación de Manning

$$Q = \frac{1}{n} A_c R^{2/3} S^{1/2}$$

donde n = coeficiente de rugosidad de Manning (número adimensional que se usa para parametrizar la fricción en el canal), A_c = área de la sección transversal del canal (m^2), S = pendiente del canal (adimensional, metros en vertical por metros en horizontal) y R = radio hidráulico (m), el cual está relacionado con otros parámetros más por $R = A_c/P$, donde P = perímetro mojado (m). Como su nombre lo dice, el perímetro mojado es la longitud de los lados y fondo del canal que están bajo el agua. Por ejemplo, para un canal rectangular, se define como $P = B + 2H$, donde H = profundidad (m). Suponga que se utiliza esta fórmula para diseñar un canal recubierto (observe que los granjeros usan canales recubiertos para minimizar las pérdidas por fugas).

a) Dados los parámetros n = 0.035, S = 0.003 y Q = 1 m^3/s, determine los valores de B y H que minimizan el perímetro mojado. Observe que dicho cálculo minimizaría el costo si los costos del recubrimiento fueran mucho mayores que los de excavación.

b) Vuelva a resolver el inciso *a*), pero incluya el costo de excavación. Para hacer esto minimice la siguiente función de costo,

$$C = c_1 A_c + c_2 P$$

donde c_1 es un factor de costo para la excavación = \$100/m^2 y c_2 es un factor de costo del recubrimiento de \$50/m.

c) Analice las implicaciones de los resultados.

16.20 Una viga cilíndrica soporta una carga de compresión de P = 3 000 kN. Para impedir que la viga se flexione (doble), la carga debe ser menor que la crítica,

$$P_c = \frac{\pi^2 EI}{L^2}$$

donde E = módulo de Young = 200 × 10^9 N/m^2, $I = \pi r^4/4$ (momento de inercia del área para una viga cilíndrica de radio r) y L es la longitud de la viga. Si el volumen de la viga V no puede exceder de 0.075 m^3, encuentre la altura más grande L que puede utilizarse, así como el radio correspondiente.

16.21 El río Splash tiene una tasa de flujo de 2 × 10^6 m^3/d, de los cuales puede derivarse hasta el 70% hacia dos canales por los que fluye a través de Splish County. Estos canales se usan para el transporte, irrigación y generación de energía eléctrica, y los últimos dos usos son fuentes de ingresos. El uso para el transporte requiere una tasa de flujo derivado mínimo de 0.3 × 10^6 m^3/d para el Canal 1 y 0.2 × 10^6 m^3/d para el Canal 2. Por razones políticas se decidió que la diferencia absoluta entre las tasas de flujo en los dos canales no excediera de 40% del flujo total derivado hacia los canales. El Organismo de Administración del Agua de Splish County, también ha

limitado los costos de mantenimiento para el sistema de canales a no más de $\$1.8 \times 10^6$ por año. Los costos anuales de mantenimiento se estiman con base en la tasa de flujo diario. Los costos por año para el Canal 1 se estiman multiplicando \$1.1 por los m^3/d de flujo; mientras que para el Canal 2 el factor de multiplicación es de \$1.4 por m^3/d. El ingreso por la generación de energía eléctrica también se estima con base en la tasa de flujo diario. Para el Canal 1 ésta es de \$4.0 por m^3/d, mientras que para el Canal 2 es de \$3.0 por m^3/d. El ingreso anual por la irrigación también se estima con base en la tasa de flujo diario, pero primero deben corregirse las tasas de flujo por las pérdidas de agua en los canales antes de que se distribuya para irrigar. Esta pérdida es de 30% en el Canal 1 y de 20% en el Canal 2. En ambos canales el ingreso es de \$3.2 por m^3/d. Determine los flujos en los canales que harían máxima la utilidad.

16.22 Determine las áreas de la sección transversal de una viga que dan como resultado el peso mínimo para la trabe que se estudió en la sección 12.2 (véase la figura 12.4). Los esfuerzos de torsión (flexión) crítica y tensión máxima de los miembros de compresión y tensión son de 10 kPsi y 20 kPsi, respectivamente. La trabe va a construirse con acero (densidad = 3.5 lb/pie-pulg^2). Observe que la longitud del miembro horizontal (2) es de 50 pies. Asimismo, recuerde que el esfuerzo en cada miembro es igual a la fuerza dividida entre el área de la sección transversal. Plantee el problema como un problema de programación lineal. Obtenga la solución en forma gráfica y con la herramienta Solver de Excel.

Ingeniería eléctrica

16.23 Alrededor de un conductor en forma de anillo de radio *a*, se encuentra una carga total *Q* distribuida uniformemente. A una distancia *x* del centro del anillo (véase la figura P16.23) se localiza una carga *q*. La fuerza que el anillo ejerce sobre la carga está dada por la ecuación

$$F = \frac{1}{4\pi e_0} \frac{qQx}{(x^2 + a^2)^{3/2}}$$

donde $e_0 = 8.85 \times 10^{-12} C^2/(N\ m^2)$, $q = Q = 2 \times 10^{-5}$ C y $a = 0.9$ m. Determine la distancia *x* donde la fuerza es máxima.

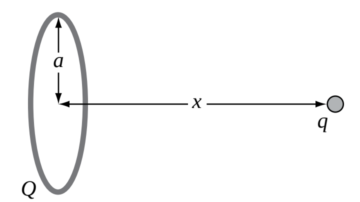


Figura P16.23

16.24 Un sistema consiste en dos plantas de energía que deben distribuir cargas por una red de transmisión. Los costos de generar la energía en las plantas 1 y 2 están dados por

$$F_1 = 2p_1 + 2$$
$$F_2 = 10p_2$$

donde p_1 y p_2 = energía producida en cada una de las plantas. Las pérdidas de energía debidas a la transmisión *L* están dadas por

$$L_1 = 0.2p_1 + 0.1p_2$$
$$L_2 = 0.2p_1 + 0.5p_2$$

La demanda total de energía es de 30 y p_1 no debe exceder de 42. Determine la generación de energía necesaria para satisfacer las demandas con el costo mínimo, con el empleo de una rutina de optimización como las que tienen, por ejemplo, Excel, software MATLAB o Mathcad.

16.25 El momento de torsión transmitido a un motor de inducción es función del deslizamiento entre la rotación del campo del estator y la velocidad del rotor *s*, donde el deslizamiento se define como

$$s = \frac{n - n_R}{n}$$

donde *n* = revoluciones por segundo de rotación de la velocidad del estator y n_R = velocidad del rotor. Pueden usarse las leyes de Kirchhoff para demostrar que el momento de torsión (expresado en forma adimensional) y el deslizamiento están relacionados por la ecuación

$$T = \frac{15(s - s^2)}{(1 - s)(4s^2 - 3s + 4)}$$

La figura P16.25 muestra esta función. Emplee un método numérico para determinar el deslizamiento con el que ocurre el momento de torsión máximo.

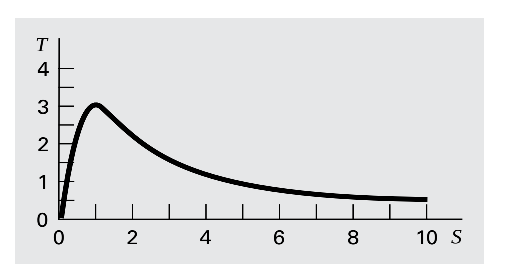


Figura P16.25
Momento de torsión transmitido a un inductor como función del deslizamiento.

16.26

a) Un fabricante de equipo de cómputo produce escáneres e impresoras. Los recursos necesarios para producirlos así como las utilidades correspondientes son los que siguen

Equipo	Capital ($/unidad)	Mano de obra (h/unidad)	Utilidad ($/unidad)
Escáner	300	20	500
Impresora	400	10	400

Si cada día se dispone de \$127 000 de capital y 4 270 horas de mano de obra, ¿qué cantidad de cada equipo debe producirse a diario a fin de maximizar la utilidad?

b) Repita el problema, pero ahora suponga que la utilidad por cada impresora vendida P_p depende del número de impresoras producidas X_p, como en

$$P_p = 400 - X_p$$

16.27 Un fabricante proporciona microcircuitos especializados. Durante los próximos tres meses, sus ventas, costos y tiempo disponible son los que siguen

	Mes 1	Mes 2	Mes 3
Circuitos requeridos	1 000	2 500	2 200
Costo del tiempo normal ($/circuito)	100	100	120
Costo del tiempo extra ($/circuito)	110	120	130
Tiempo de operación regular (h)	2 400	2 400	2 400
Tiempo extra (h)	720	720	720

Al principio del primer mes no existen circuitos almacenados. Toma 1.5 horas del tiempo de producción fabricar un circuito y cuesta $5 almacenarlo de un mes al siguiente. Determine un programa de producción que satisfaga los requerimientos de la demanda, sin que exceda las restricciones de tiempo de producción mensual, y minimice el costo. Observe que al final de los 3 meses no debe haber circuitos almacenados.

Ingeniería mecánica/aeroespacial

16.28 El arrastre total de un aeroplano se estima por medio de

$$D = 0.01\sigma V^2 + \frac{0.95}{\sigma}\left(\frac{W}{V}\right)^2$$

fricción empuje

donde D = arrastre, σ = razón de la densidad del aire entre la altitud de vuelo y el nivel del mar, W = peso y V = velocidad. Como se observa en la figura P16.28, los dos factores que contribuyen al arrastre resultan afectados en forma distinta conforme la velocidad aumenta. Mientras que el arrastre por fricción se incrementa con la velocidad, el arrastre debido al empuje disminuye. La combinación de los dos factores lleva a un arrastre mínimo.

a) Si $\sigma = 0.5$ y $W = 15\,000$, determine el arrastre mínimo y la velocidad a la que ocurre.

b) Además, realice un análisis de sensibilidad para determinar cómo varía este óptimo en respuesta a un rango de $W = 12\,000$ a $18\,000$ con $\sigma = 0.5$.

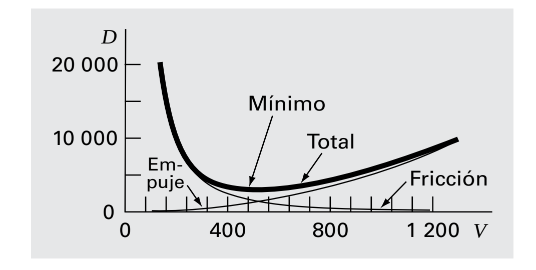


Figura P16.28
Gráfica de arrastre *versus* la velocidad de un aeroplano.

16.29 Los baleros de rodamiento están expuestos a fallar por la fatiga ocasionada por cargas grandes de contacto F (véase la figura P16.29).

Puede demostrarse que el problema de encontrar la ubicación del esfuerzo máximo a lo largo del eje x es equivalente a maximizar la función

$$f(x) = \frac{0.4}{\sqrt{1+x^2}} - \sqrt{1+x^2}\left(1 - \frac{0.4}{1+x^2}\right) + x$$

Encuentre el valor de x que maximiza a $f(x)$.

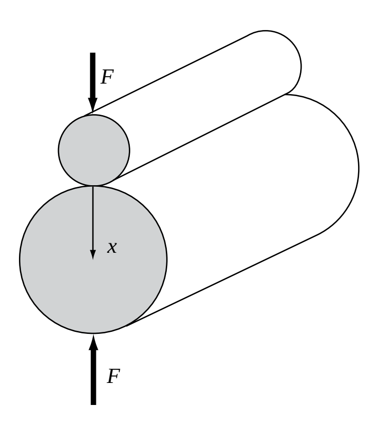


Figura P16.29
Baleros de rodamiento.

16.30 Una compañía aeroespacial desarrolla un aditivo nuevo para el combustible de aeronaves comerciales. El aditivo está compuesto de tres ingredientes: X, Y y Z. Para el rendimiento mayor, la cantidad total de aditivo debe ser al menos de 6 mL/L de combustible. Por razones de seguridad, la suma de los ingredientes X y Y altamente inflamables, no debe exceder los 2.5 mL/L. Además, la cantidad del ingrediente X siempre debe ser mayor o igual a la de Y, y la de Z debe ser mayor que la mitad de la de Y. Si el costo por mL para los ingredientes X, Y y Z es de 0.05, 0.025 y 0.15, respectivamente, determine la mezcla de costo mínimo para un litro de combustible.

16.31 Una empresa manufacturera produce cuatro tipos de partes automotrices. Cada una de ellas primero se fabrica y luego se le dan los acabados. Las horas de trabajador requeridas y la utilidad para cada parte son las siguientes

	Parte			
	A	B	C	D
Tiempo de fabricación (h/100 unidades)	2.5	1.5	2.75	2
Tiempo de acabados (h/100 unidades)	3.5	3	3	2
Utilidad ($/100 unidades)	375	275	475	325

Las capacidades de los talleres de fabricación y acabados para el mes siguiente son de 640 y 960 horas, respectivamente. Determine qué cantidad de cada parte debe producirse a fin de maximizar la utilidad.

16.32 De modo similar al estudio de caso que se describe en la sección 16.4, desarrolle la función de energía potencial para el sistema que se ilustra en la figura P16.32. Desarrolle gráficas de contorno y superficie en MATLAB. Minimice la función de energía potencial con el fin de determinar los desplazamientos en el equilibrio x_1 y x_2 dada la función de forzamiento F = 100 N, y el parámetro k_a = 20 y k_b = 15 N/m.

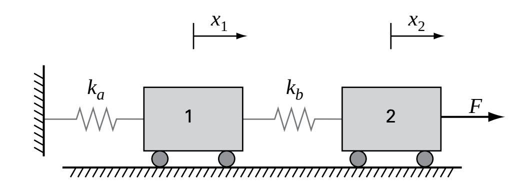


Figura P16.32
Dos masas sin fricción conectadas a una pared por un par de resortes elásticos lineales.

16.33 El reciente interés en el ciclismo de competencia y recreativo ha dado como resultado que los ingenieros dirigieran sus habilidades hacia el diseño y prueba de bicicletas (figura P16.33*a*). Suponga que se le ha asignado la tarea de predecir el desplazamiento horizontal y vertical de un sistema de armazón de bicicleta en respuesta a una fuerza. Suponga que las fuerzas que usted debe analizar se pueden simplificar como se ilustra en la figura P16.33*b*. Usted está interesado en probar la respuesta del tirante a una fuerza que se ejerza en cualquier número de direcciones designadas por el ángulo θ. Los parámetros para el problema son E = módulo de Young = 2×10^{11} Pa, A = área de la sección transversal = 0.0001 m^2, w = anchura = 0.44 m, ℓ = longitud = 0.56 m y h = altura = 0.5 m. Los desplazamientos x y y se pueden resolver determinando los valores que dan una energía potencial mínima. Determine los desplazamientos para una fuerza de 10 000 N y un rango de θ de 0° (horizontal) a 90° (vertical).

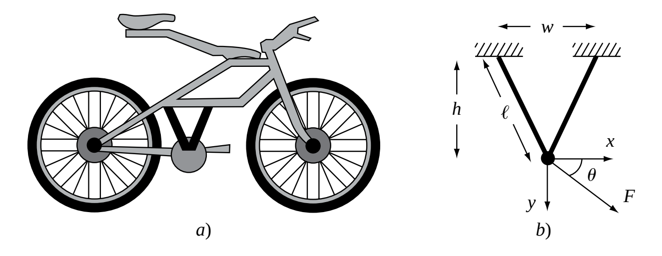


Figura P16.33
a) Una bicicleta con *b*) un diagrama de cuerpo libre para una parte del marco.

EPÍLOGO: PARTE CUATRO

Los epílogos de las otras partes de este libro contienen un análisis y un resumen tabular de las ventajas y desventajas de los diferentes métodos, así como las fórmulas y relaciones importantes. La mayoría de los métodos de esta parte son complicados y, en consecuencia, no se pueden resumir en fórmulas simples y tablas. Por lo tanto, aquí nos desviaremos un poco para ofrecer el siguiente análisis escrito de las alternativas y las referencias adicionales.

PT4.4 ALTERNATIVAS

En el capítulo 13 se trató de la búsqueda del valor óptimo de una función con una sola variable sin restricciones. El método de búsqueda de la sección dorada es un método cerrado que requiere de un intervalo que contenga un solo valor óptimo conocido. Tiene la ventaja de minimizar las evaluaciones de la función, y ser siempre convergente. La interpolación parabólica funciona mejor cuando se implementa como un método cerrado, aunque también se puede programar como un método abierto. Sin embargo, en tales casos, puede diverger. Tanto el método de búsqueda de la sección dorada como el de interpolación parabólica no requieren evaluaciones de la derivada. Así, ambos son apropiados cuando el intervalo puede definirse fácilmente y las evaluaciones de la función son demasiadas.

El método de Newton es un método abierto que no requiere que esté dentro de un intervalo óptimo. Puede implementarse en una representación de forma cerrada, cuando la primera y segunda derivadas se determinan en forma analítica. También se implementa en una forma similar el método de la secante al representar las derivadas en diferencias finitas. Aunque el método de Newton converge rápidamente cerca del óptimo, puede diverger con valores iniciales pobres. Además la convergencia depende también de la naturaleza de la función.

Finalmente, hay enfoques híbridos disponibles que instrumentan diversos métodos para obtener tanto confiabilidad como eficiencia. El método de Brent hace esto combinando la confiable búsqueda de sección dorada con la rápida interpolación parabólica.

En el capítulo 14 se trataron dos tipos generales de métodos para resolver problemas de optimización sin restricciones multidimensionales. Los métodos directos como el de búsquedas aleatorias y el de búsquedas univariables no requieren el cálculo de las derivadas de la función y con frecuencia son ineficientes. Sin embargo, proporcionan también una herramienta para encontrar el óptimo global más que el local. Los métodos de búsqueda con un patrón como el método de Powell llegan a ser muy eficientes y tampoco requieren del cálculo de la derivada.

Los métodos con gradiente usan la primera y, algunas veces, la segunda derivadas para encontrar el óptimo. El método del mayor ascenso/descenso ofrece un procedimiento confiable pero en ocasiones lento. Por el contrario, el método de Newton converge con rapidez cuando se está en la vecindad de una raíz; pero algunas veces sufre de divergencia. El método de Marquardt utiliza el método de mayor descenso en la ubicación inicial, muy lejos del óptimo, y después cambia al método de Newton cerca del óptimo, en un intento por aprovechar las fortalezas de cada método.

El método de Newton puede ser costoso computacionalmente ya que requiere calcular tanto del vector gradiente como de la matriz hessiana. Los métodos cuasi-Newton intentan evitar estos problemas al usar aproximaciones para reducir el número de evaluaciones de matrices (particularmente, la evaluación, el almacenamiento y la inversión del hessiano).

En la actualidad, las investigaciones continúan para explorar las características y las ventajas correspondientes de varios métodos híbridos y en tándem. Algunos ejemplos son el método del gradiente conjugado de Fletcher-Reeves y los métodos cuasi-Newton de Davidon-Fletcher-Powell.

El capítulo 15 se dedicó a la optimización con restricciones. Para problemas lineales, la programación lineal basada en el método simplex ofrece un medio eficiente para obtener soluciones. Procedimientos como el método GRG sirven para resolver problemas con restricciones no lineales.

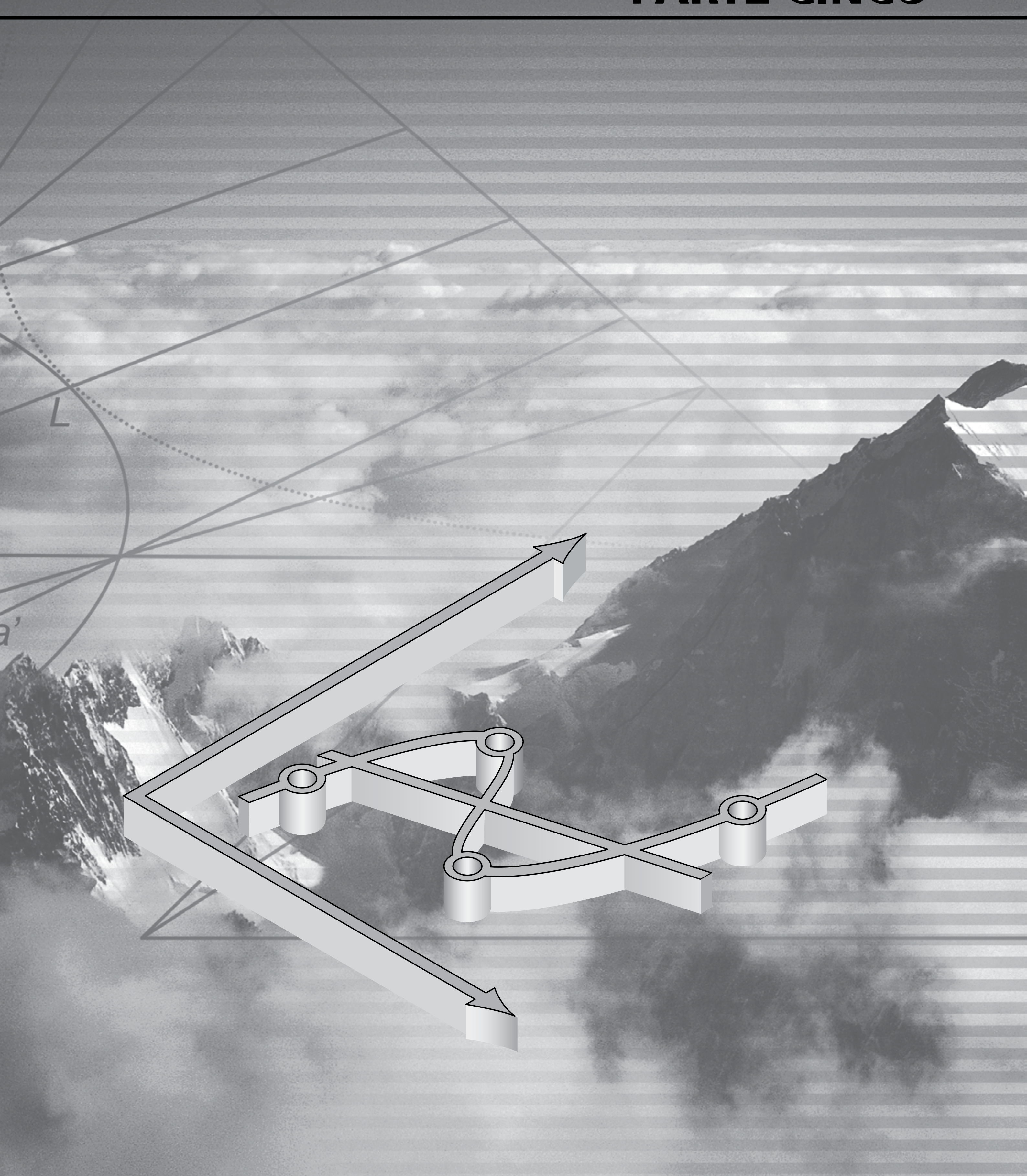
Los paquetes de software incluyen una amplia variedad de capacidades de optimización. Como se describe en el capítulo 15, Excel, MATLAB y Mathcad tienen, todos, capacidades interconstruidas de búsqueda que se pueden usar para los problemas tanto unidimensionales como bidimensionales que se encuentran rutinariamente en la ingeniería y en la ciencia.

PT4.5 REFERENCIAS ADICIONALES

En Press *et al*. (2007) y en Moler (2004) se pueden encontrar descripciones generales de optimización, incluyendo algunos algoritmos. Para problemas multidimensionales, se puede encontrar información adicional en Dennis y Schnabel (1996), Fletcher (1980, 1981), Gill *et al*. (1981) y Luenberger (1984).

Además, hay varios métodos avanzados que son adecuados para contextos específicos de problemas. Por ejemplo, los algoritmos genéticos usan estrategias inspiradas en la biología evolutiva, como la herencia, la mutación y la selección. Como no hacen suposiciones respecto al espacio de búsqueda, tales algoritmos evolutivos resultan a menudo útiles para grandes problemas con muchos locales óptimos. Las técnicas relacionadas con esto incluyen el recocido simulado y la búsqueda tabú. Hillier y Lieberman (2005) dan descripciones generales de éstas y muchas otras técnicas avanzadas.

PARTE CINCO



AJUSTE DE CURVAS

PT5.1 MOTIVACIÓN

Es común que los datos se den como valores discretos a lo largo de un continuo. Sin embargo, quizás usted requiera la estimación de un punto entre valores discretos. Esta parte del libro describe las técnicas para ajustar curvas a estos datos para obtener estimaciones intermedias. Además, usted puede necesitar la versión simplificada de una función complicada. Una manera de hacerlo es calcular valores de la función en un número discreto de valores en el intervalo de interés. Después, se obtiene una función más simple para ajustar dichos valores. Estas dos aplicaciones se conocen como *ajuste de curvas.*

Existen dos métodos generales para el ajuste de curvas que se distinguen entre sí al considerar la cantidad de error asociado con los datos. Primero, si los datos exhiben un grado significativo de error o “ruido”, la estrategia será obtener una sola curva que represente la tendencia general de los datos. Como cualquier dato individual puede ser incorrecto, no se busca intersecar todos los puntos. En lugar de esto, se construye una curva que siga la tendencia de los puntos tomados como un grupo. Un procedimiento de este tipo se llama *regresión por mínimos cuadrados* (figura PT5.1*a*).

Segundo, si se sabe que los datos son muy precisos, el procedimiento básico será colocar una curva o una serie de curvas que pasen por cada uno de los puntos en forma directa. Usualmente tales datos provienen de tablas. Como ejemplos se tienen los valores de la densidad del agua o la capacidad calorífica de los gases en función de la temperatura. La estimación de valores entre puntos discretos bien conocidos se llama *interpolación* (figuras PT5.1*b* y PT5.1*c*).

PT5.1.1 Métodos sin computadora para el ajuste de curvas

El método más simple para ajustar una curva a los datos consiste en ubicar los puntos y después trazar una curva que visualmente se acerque a los datos. Aunque ésta es una operación válida cuando se requiere una estimación rápida, los resultados dependen del punto de vista subjetivo de la persona que dibuja la curva.

Por ejemplo, en la figura PT5.1 se muestran curvas trazadas a partir del mismo conjunto de datos por tres ingenieros. El primero no intentó unir los puntos, sino, más bien, caracterizar la tendencia general ascendente de los datos con una línea recta (figura PT5.1*a*). El segundo ingeniero usó segmentos de línea recta o interpolación lineal para unir los puntos (figura PT5.1*b*). Ésta es una práctica común en la ingeniería. Si los valores se encuentran cercanos a ser lineales o están cercanamente espaciados, tal aproximación ofrece estimaciones que son adecuadas en muchos cálculos de ingeniería. No obstante, si la relación es altamente curvilínea o los datos están muy espaciados, es posible introducir errores mediante esa interpolación lineal. El tercer ingeniero utiliza curvas suaves para tratar de capturar el serpenteado sugerido por los datos (figura PT5.1*c*). Un cuarto o quinto ingeniero podría, de igual forma, desarrollar ajustes alternativos. Obviamente, nuestra meta aquí es desarrollar métodos sistemáticos y objetivos con el propósito de obtener tales curvas.

PT5.1.2 Ajuste de curvas y práctica en ingeniería

Su primer encuentro con el ajuste de curvas podría haber sido determinar valores intermedios a partir de datos tabulados (por ejemplo, tablas de interés para ingeniería económica, o tablas de vapor en termodinámica). En lo que resta de su carrera, usted tendrá frecuentes oportunidades para estimar valores intermedios a partir de tablas.

Aunque se han tabulado muchas propiedades ampliamente utilizadas en la ingeniería, existen otras que no están disponibles en esta forma conveniente. Los casos especiales y nuevos contextos de problemas requieren que usted recolecte sus propios datos y desarrolle sus propias relaciones

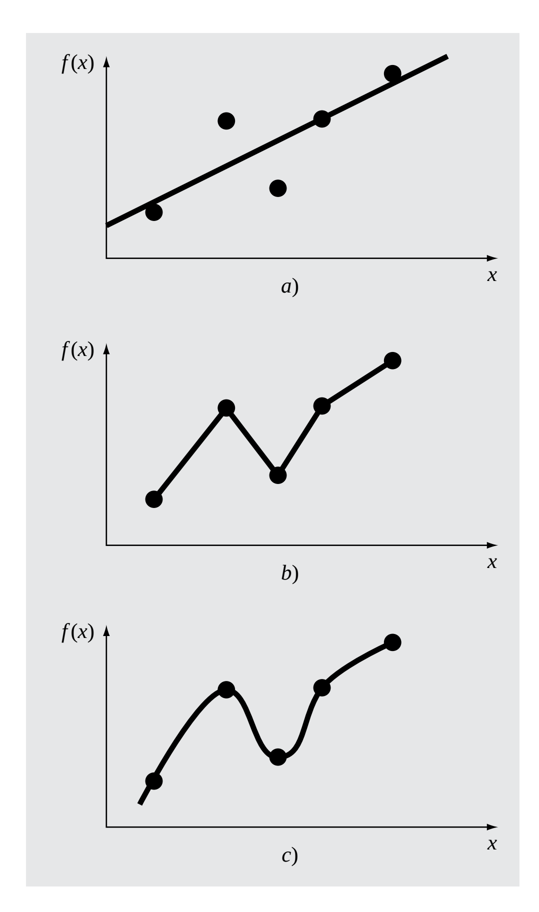


Figura PT5.1
Tres intentos para ajustar una “mejor” curva con cinco puntos dados. *a*) Regresión por mínimos cuadrados, *b*) interpolación lineal y *c*) interpolación curvilínea.

predictivas. Se han encontrado dos tipos de aplicaciones en el ajuste de datos experimentales: análisis de la tendencia y prueba de hipótesis.

El análisis de la tendencia representa el proceso de utilizar el comportamiento de los datos para realizar predicciones. En casos donde los datos son medidas de alta precisión, se usan polinomios de interpolación. Los datos imprecisos se analizan mediante una regresión por mínimos cuadrados.

El *análisis de la tendencia* sirve para predecir o pronosticar valores de la variable dependiente. Esto puede implicar una extrapolación más allá de los límites de los datos observados o una interpolación dentro del intervalo de los datos. Por lo común, en todos los campos de la ingeniería se presentan problemas de este tipo.

Una segunda aplicación del ajuste de curvas experimental en ingeniería es la *prueba de hipótesis*. Aquí, un modelo matemático existente se compara con los datos obtenidos. Si se desconocen los coeficientes del modelo, será necesario determinar los valores que mejor se ajusten a los datos observados. Por otro lado, si ya se dispone de la estimación de los coeficientes del modelo sería conveniente comparar los valores predichos del modelo con los observados para probar qué tan adecuado es el modelo. Con frecuencia, se comparan modelos alternativos y se elige “el mejor” considerando las observaciones hechas en forma empírica.

Además de las aplicaciones mencionadas en la ingeniería, el ajuste de curvas es importante para implementar otros métodos numéricos, como la integración y la solución aproximada de ecuaciones diferenciales. Por último, las técnicas de ajuste de curvas son útiles para obtener funciones simples con la finalidad de aproximar funciones complicadas.

PT5.2 ANTECEDENTES MATEMÁTICOS

Los fundamentos matemáticos de la interpolación se encuentran en el conocimiento sobre las expansiones de la serie de Taylor y las diferencias finitas divididas que se presentaron en el capítulo 4. La regresión por mínimos cuadrados requiere además de la información en el campo de la estadística. Si usted conoce los conceptos de la media, desviación estándar, suma del cuadrado de los residuales, distribución normal e intervalos de confianza, puede omitir el estudio de las siguientes páginas y pasar directamente a la sección PT5.3. Si no recuerda muy bien estos conceptos o necesita de un repaso, el estudio del siguiente material le servirá como introducción a esos temas.

PT5.2.1 Estadística simple

Suponga que en el curso de un estudio de ingeniería se realizaron varias mediciones de una cantidad específica. Por ejemplo, la tabla PT5.1 contiene 24 lecturas del coeficiente de expansión térmica del acero. Tomados así, los datos ofrecen una información limitada (es decir, que los valores tienen un mínimo de 6.395 y un máximo de 6.775). Se obtiene una mayor comprensión al analizar los datos mediante uno o más estadísticos, bien seleccionados, que den tanta información como sea posible acerca de las características específicas del conjunto de datos. Esos estadísticos descriptivos se seleccionan para representar 1) la posición del centro de la distribución de los datos y 2) el grado de dispersión de los datos.

El estadístico de posición más común es la media aritmética. La *media aritmética* ($\bar{y}$) de una muestra se define como la suma de los datos (y_i) dividida entre el número de datos (n), o

$$\bar{y} = \frac{\Sigma y_i}{n} \qquad \text{(PT5.1)}$$

donde la sumatoria (y todas las sumatorias que siguen en esta introducción) va desde i = 1 hasta n.

La medida de dispersión más común para una muestra es la *desviación estándar* (s_y) respecto de la media,

$$s_y = \sqrt{\frac{S_t}{n-1}} \tag{PT5.2}$$

donde S_t es la suma total de los cuadrados de las diferencias entre los datos y la media, o

$$S_t = \Sigma(y_i - \bar{y})^2 \tag{PT5.3}$$

Así, si las mediciones se encuentran muy dispersas alrededor de la media, S_t (y, en consecuencia, s_y) será grande. Si están agrupadas cerca de ella, la desviación estándar será pequeña. La dispersión también se puede representar por el cuadrado de la desviación estándar, llamada la *varianza*:

$$s_y^2 = \frac{\Sigma(y_i - \bar{y})^2}{n-1} \tag{PT5.4}$$

Observe que el denominador en ambas ecuaciones (PT5.2) y (PT5.4) es n – 1. La cantidad n – 1 se conoce como los grados de libertad. Por lo tanto, se dice que S_t y s_y consideran n – 1 *grados de libertad*. Esta nomenclatura se obtiene del hecho de que la suma de las cantidades sobre las cuales se basa S_t (es decir, $\bar{y} - y_1, \bar{y} - y_2, \ldots, \bar{y} - y_n$) es cero. En consecuencia, si se conoce $\bar{y}$ y se especifican los valores de n – 1, el valor restante queda determinado. Así, sólo n – 1 de los valores se dice que se determinan libremente. Otra justificación para dividir entre n – 1 es el hecho de que no tiene sentido hablar de la dispersión de un solo dato. Cuando n = 1, las ecuaciones (PT5.2) y (PT5.4) dan un resultado sin sentido: infinito.

Se deberá observar que hay otra fórmula alternativa más conveniente, para calcular la desviación estándar,

$$s_y^2 = \frac{\Sigma y_i^2 - (\Sigma y_i)^2/n}{n-1}$$

Esta versión no requiere el cálculo previo de $\bar{y}$ y se obtiene el mismo resultado que con la ecuación (PT5.4).

Un estadístico final que tiene utilidad para cuantificar la dispersión de los datos es el *coeficiente de variación* (*c.v.*). Tal estadístico es el cociente de la desviación estándar entre la media. De esta manera, proporciona una medición normalizada de la dispersión. Con frecuencia se multiplica por 100 para expresarlo como porcentaje:

$$\text{c.v.} = \frac{s_y}{\bar{y}} 100\% \tag{PT5.5}$$

Observe que el coeficiente de variación tiene un carácter similar al del error relativo porcentual (ε_t) analizado en la sección 3.3. Es decir, éste es la razón de una medición de error (s_y) respecto a un estimado del valor verdadero ($\bar{y}$).

Tabla PT5.1 Mediciones del coeficiente de expansión térmica del acero [$\times 10^{-6}$ in/(in · °F)].

6.495	6.595	6.615	6.635	6.485	6.555
6.665	6.505	6.435	6.625	6.715	6.655
6.755	6.625	6.715	6.575	6.655	6.605
6.565	6.515	6.555	6.395	6.775	6.685

EJEMPLO PT5.1 Estadística simple de una muestra

Planteamiento del problema Calcule la media, la varianza, la desviación estándar y el coeficiente de variación para los datos de la tabla PT5.1.

Solución Se suman los datos (tabla PT5.2) y los resultados sirven para calcular [ecuación (PT5.1)]

$$\bar{y} = \frac{158.4}{24} = 6.6$$

Como se observa en la tabla PT5.2, la suma de los cuadrados de las diferencias es 0.217000, los cuales se usan para calcular la desviación estándar [ecuación (PT5.2)]:

$$s_y = \sqrt{\frac{0.217000}{24-1}} = 0.097133$$

la varianza [ecuación (PT5.4)]:

$$s_y^2 = 0.009435$$

Tabla PT5.2 Cálculos para estadísticos con las lecturas del coeficiente de expansión térmica. Las frecuencias y los límites se calculan para construir el histograma que se muestra en la figura PT5.2.

				Intervalo	
i	y_i	$(y_i - \bar{y})^2$	Frecuencia	Límite inferior	Límite superior
1	6.395	0.042025	1	6.36	6.40
2	6.435	0.027225	1	6.40	6.44
3	6.485	0.013225			
4	6.495	0.011025			
5	6.505	0.009025	4	6.48	6.52
6	6.515	0.007225			
7	6.555	0.002025	2	6.52	6.56
8	6.555	0.002025			
9	6.565	0.001225			
10	6.575	0.000625	3	6.56	6.60
11	6.595	0.000025			
12	6.605	0.000025			
13	6.615	0.000225			
14	6.625	0.000625	5	6.60	6.64
15	6.625	0.000625			
16	6.635	0.001225			
17	6.655	0.003025			
18	6.655	0.003025	3	6.64	6.68
19	6.665	0.004225			
20	6.685	0.007225			
21	6.715	0.013225	3	6.68	6.72
22	6.715	0.013225			
23	6.755	0.024025	1	6.72	6.76
24	6.775	0.030625	1	6.76	6.80
Σ	158.4	0.217000			

y el coeficiente de variación [ecuación (PT5.5)]:

$$\text{c.v.} = \frac{0.097133}{6.6} 100\% = 1.47\%$$

PT5.2.2 La distribución normal

Otra característica útil en el presente análisis es la *distribución de datos* (es decir, la forma en que los datos se distribuyen alrededor de la media). Un *histograma* proporciona una representación visual simple de la distribución. Como se observa en la tabla PT5.2, el histograma se construye al ordenar las mediciones en intervalos. Las unidades de medición se grafican en las abscisas; y la frecuencia de ocurrencia de cada intervalo, en las ordenadas. Así, cinco de las mediciones se encuentran entre 6.60 y 6.64. Como se advierte en la figura PT5.2, el histograma indica que la mayoría de los datos se agrupa cerca del valor de la media de 6.6.

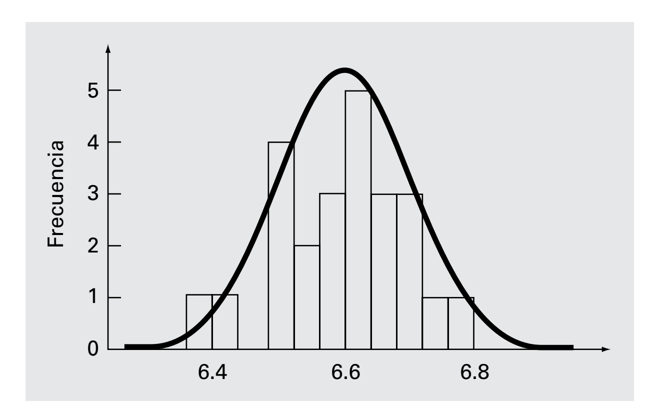


Figura PT5.2
Histograma usado para ilustrar la distribución de datos. Conforme el número de datos aumenta, el histograma se aproximará a una curva suave, la curva en forma de campana, llamada la distribución normal.

Si se tiene un conjunto muy grande de datos, el histograma se puede aproximar mediante una curva suave. La curva simétrica, en forma de campana que se sobrepone en la figura PT5.2, es una de estas formas características (la *distribución normal*). Dadas suficientes mediciones, en este caso particular el histograma se aproximará a la distribución normal.

Los conceptos de media, desviación estándar, suma del cuadrado de los residuales y distribución normal tienen una gran importancia en la práctica de la ingeniería. Un ejemplo muy simple es su uso para cuantificar la confianza que se puede tener en una medición en particular. Si una cantidad está normalmente distribuida, el intervalo limitado por $\bar{y} - s_y$ y $\bar{y} + s_y$ abarcará en forma aproximada el 68% de las mediciones totales. De manera similar, el intervalo limitado por $\bar{y} - 2s_y$ y $\bar{y} + 2s_y$ abarcará alrededor del 95%.

Por ejemplo, para los datos de la tabla PT5.1 ($\bar{y} = 6.6$ y $s_y = 0.097133$), se afirma que aproximadamente el 95% de las lecturas deberán estar entre 6.405734 y 6.794266. Si alguien nos dijera que tomó una lectura de 7.35, entonces sospecharíamos que la medición resultó errónea. En la siguiente sección se estudiarán dichas evaluaciones.

PT5.2.3 Estimación de los intervalos de confianza

Como resultará claro de lo expuesto en la sección anterior, uno de los principales objetivos de la estadística es estimar las propiedades de una *población* basándose en una *muestra* limitada que se toma de esa población. Es evidente que es imposible medir el coeficiente de expansión térmica de cada pieza producida de acero. En consecuencia, como se muestra en las tablas PT5.1 y PT5.2, es posible realizar un número de mediciones en forma aleatoria y, con base en la muestra, intentar caracterizar las propiedades de toda la población.

Debido a que se "infieren" propiedades de la población desconocida a partir de una muestra limitada, el procedimiento se denomina *inferencia estadística*. Ya que los resultados a menudo se reportan como estimaciones de los parámetros de la población, el proceso también se conoce como *estimación*.

Ya se mostró cómo estimar la tendencia central (media de la muestra, $\bar{y}$) y la dispersión (desviación estándar y varianza de la muestra) de una muestra limitada. Ahora, se describirá en forma breve cómo realizar aseveraciones probabilísticas respecto de la calidad de esas estimaciones. En particular, se analizará cómo definir un intervalo de confianza alrededor de un estimado de la media. Se ha escogido este tópico en particular debido a su relevancia directa para los modelos de regresión que se describirán en el capítulo 17.

En el siguiente análisis observe que la nomenclatura $\bar{y}$ y s_y se refieren a la media de la muestra y a su desviación estándar, respectivamente. La nomenclatura μ y σ se refieren a la media y la desviación estándar de la población. Las primeras son algunas veces referidas como la media y desviación estándar "estimadas"; mientras que las últimas se llaman la media y la desviación estándar "verdaderas".

Un *estimador de intervalo* proporciona el rango de valores dentro del que se espera que esté el parámetro, con una probabilidad dada. Tales intervalos se describen como unilateral y bilateral.

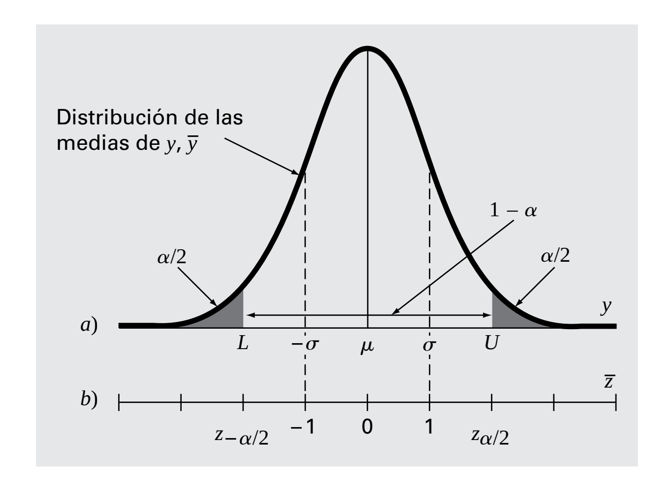


Figura PT5.3
Un intervalo de confianza bilateral. La escala de la abscisa en *a*) se escribe en las unidades originales de la variable aleatoria *y*. *b*) Es una versión normalizada de las abscisas que tiene la media ubicada en el origen y se escala el eje de tal manera que la desviación estándar corresponda a una unidad.

Como su nombre lo indica, un *intervalo unilateral* expresa nuestra confianza en que el parámetro estimado sea menor que o mayor que el valor real. En cambio, el *intervalo bilateral* tiene que ver con la proposición más general en que la estimación concuerda con la verdad, sin considerar el signo de la discrepancia. Como éste es más general, nos ocuparemos del intervalo bilateral.

Un intervalo bilateral se describe con la relación

$$P\{L \le \mu \le U\} = 1 - \alpha$$

que se lee: "la probabilidad de que la media real de y, μ, esté dentro de los límites de L a U es $1 - \alpha$". La cantidad α se conoce como el *nivel de significancia*. De esta forma, el problema de definir un intervalo de confianza se reduce a estimar L y U. Aunque no es absolutamente necesario, es costumbre visualizar el intervalo bilateral con la probabilidad α, distribuida de manera uniforme, con $\alpha/2$ en cada cola de la distribución, como se muestra en la figura PT5.3.

Si se conoce la varianza real de la distribución de y, σ^2 (lo cual no es frecuente), la teoría estadística establece que la media de la muestra $\bar{y}$ proviene de una distribución normal con media μ y varianza σ^2/n (cuadro PT5.1). En el caso ilustrado en la figura PT5.3, no se conoce realmente μ. Por lo tanto, no se sabe dónde se ubica con exactitud la curva normal con respecto a $\bar{y}$. Para evitar este dilema, se calcula una nueva cantidad, el *estimador normal estándar*

Cuadro PT5.1 Un poco de estadística

La mayoría de los ingenieros toman varios cursos de estadística. Como usted tal vez aún no ha tomado alguno se mencionarán algunas nociones que harán que esta sección sea más coherente.

Como se ha mencionado, el "juego" de la estadística inferencial supone que la variable aleatoria que usted muestrea, y, tiene media (μ) y varianza (σ^2) verdaderas. Además, en este análisis se supondrá que tiene una distribución particular: la distribución normal. La varianza de esta distribución normal tiene un valor finito que especifica la "dispersión" de la distribución normal. Si la varianza es grande, la distribución es amplia. En cambio, si la varianza es pequeña, la distribución es estrecha. Así, la varianza real cuantifica la incertidumbre intrínseca de la variable aleatoria.

En el juego de la estadística, se toma un número limitado de mediciones de una cantidad, a la que se le llama muestra. De esta muestra, se calculan una media ($\bar{y}$) y una varianza (s_y^2) estimadas. Cuantas más mediciones se tomen, mejor serán las estimaciones para que se aproximen a los valores verdaderos. Esto es, cuando $n \to \infty$, $\bar{y} \to \mu$ y $s_y^2 \to \sigma^2$.

Suponga que se toman n muestras y se calcula una media estimada $\bar{y}_1$. Después se toman otras n muestras y se calcula otra, $\bar{y}_2$. Se puede repetir este proceso hasta que se haya generado una muestra de medias: $\bar{y}_1, \bar{y}_2, \bar{y}_3, \ldots, \bar{y}_m$, donde m es grande. Entonces se construye un histograma de estas medias y se determina una "distribución de las medias", así como una "media de las medias" y una "desviación estándar de las medias". Ahora surge la pregunta: ¿esta nueva distribución de medias y sus estadísticos se comportan en una forma predecible?

Existe un teorema muy importante conocido como el *teorema del límite central* que responde en forma directa a esta pregunta y se enuncia como sigue

Sea $y_1, y_2, \ldots, y_n$, una muestra aleatoria de tamaño n tomada de una distribución con media μ y varianza σ^2. Entonces, para n grandes, $\bar{y}$ es aproximadamente normal con la media μ y la varianza σ^2/n. Además, para n grande, la variable aleatoria $(\bar{y} - \mu)/(\sigma/\sqrt{n})$ es aproximadamente normal estándar.

Así, el teorema establece el resultado interesante de que la distribución de las medias siempre estará normalmente distribuida, ¡sin importar la distribución de las variables aleatorias de que se trate! Esto también da el resultado esperado, de que dada una muestra suficientemente grande, la media de las medias deberá converger hacia la verdadera media de la población μ.

Además, el teorema indica que conforme crezca el tamaño de la muestra, la varianza de las medias se aproximará a cero. Esto tiene sentido, ya que si n es pequeña, las estimaciones individuales de la media serán pobres, y las varianzas de las medias, grandes. En tanto n aumente, la estimación de la media mejorará y, por lo tanto, disminuirá su dispersión. El teorema del límite central claramente define, en forma exacta, cómo esta disminución está relacionada tanto con la varianza real como con el tamaño de la muestra; es decir, como σ^2/n.

Por último, el teorema establece el importante resultado que se ha dado en la ecuación (PT5.6). Como se muestra en esta sección, este teorema es la base para construir intervalos de confianza para la media.

$$\bar{z} = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} \tag{PT5.6}$$

que representa la distancia normalizada entre $\bar{y}$ y μ. De acuerdo con la teoría estadística, esta cantidad deberá estar distribuida normalmente con media 0 y varianza 1. Además, la probabilidad de que $\bar{z}$ esté dentro de la región no sombreada de la figura PT5.3 será $1 - \alpha$. Por lo tanto, se establece que

$$\frac{\bar{y} - \mu}{\sigma/\sqrt{n}} < -z_{\alpha/2} \qquad \text{o} \qquad \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} > z_{\alpha/2}$$

con una probabilidad de α.

La cantidad $z_{\alpha/2}$ es una variable aleatoria normal estándar. Ésta es la distancia medida a lo largo del eje normalizado arriba y debajo de la media, que corresponde la probabilidad $1 - \alpha$ (figura PT5.3*b*). Los valores de $z_{\alpha/2}$ están tabulados en libros de estadística (por ejemplo, Milton y Arnold, 2002). También pueden calcularse usando funciones de paquetes de software como Excel, MATLAB y Mathcad. Como un ejemplo, para $\alpha = 0.05$ (en otras palabras, definiendo un intervalo que comprenda 95%), $z_{\alpha/2}$ es aproximadamente igual a 1.96. Esto significa que un intervalo alrededor de la media con un ancho ±1.96 veces la desviación estándar abarcará, en forma aproximada, el 95% de la distribución.

Esos resultados se reordenan para obtener

$$L \le \mu \le U$$

con una probabilidad de $1 - \alpha$, donde

$$L = \bar{y} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \qquad U = \bar{y} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \tag{PT5.7}$$

Ahora, aunque lo anterior ofrece una estimación de L y U, está basado en el conocimiento de la verdadera varianza σ. Y en nuestro caso, conocemos solamente la varianza estimada s_y. Una alternativa inmediata sería una versión de la ecuación (PT5.6) basada en s_y:

$$t = \frac{\bar{y} - \mu}{s_y/\sqrt{n}} \tag{PT5.8}$$

Aun cuando la muestra se tome de una distribución normal, esta fracción no estará normalmente distribuida, en particular cuando n sea pequeña. W. S. Gossett encontró que la variable aleatoria definida por la ecuación (PT5.8) sigue la llamada distribución t de Student o, simplemente, *distribución t*. En este caso,

$$L = \bar{y} - \frac{s_y}{\sqrt{n}} t_{\alpha/2,n-1} \qquad U = \bar{y} + \frac{s_y}{\sqrt{n}} t_{\alpha/2,n-1} \tag{PT5.9}$$

donde $t_{\alpha/2,\,n-1}$ es la variable aleatoria estándar de la distribución t para una probabilidad de $\alpha/2$. Como en el caso de $z_{\alpha/2}$, los valores están tabulados en libros de estadística, y también se calculan mediante paquetes de software. Por ejemplo, si $\alpha = 0.05$ y $n = 20$, $t_{\alpha/2,\,n-1} = 2.086$.

La distribución t puede entenderse como una modificación de la distribución normal que toma en cuenta el hecho de que se tiene una estimación imperfecta de la desviación estándar. Cuando n es pequeña, tiende a ser más plana que la normal (figura PT5.4). Entonces, para pocas mediciones, se obtienen intervalos de confianza más amplios y, por lo tanto, más conservadores. Conforme n se vuelve más grande, la distribución t converge a la normal.

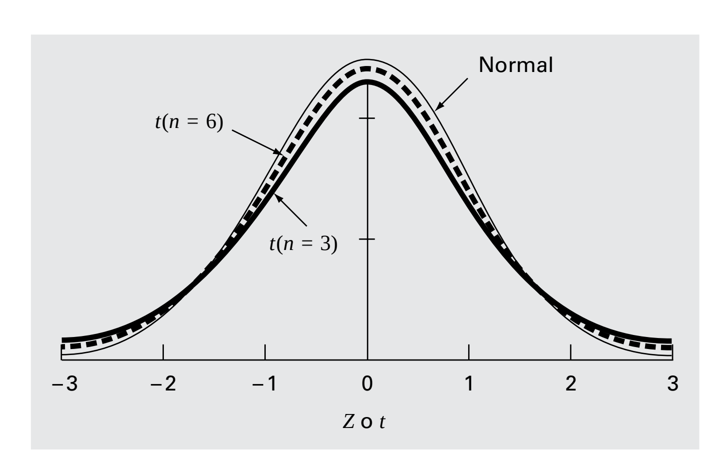


Figura PT5.4
Comparación de la distribución normal con la distribución t para $n = 3$ y $n = 6$. Observe cómo la distribución t en general es más plana.

EJEMPLO PT5.2 Intervalo de confianza alrededor de la media

Planteamiento del problema Determine la media y el correspondiente intervalo de confianza del 95% para los datos de la tabla PT5.1. Realice 3 estimaciones basándose en *a*) las primeras 8 mediciones, *b*) las primeras 16 mediciones y *c*) las 24 mediciones.

Solución *a*) La media y la desviación estándar con los primeros 8 valores es

$$\bar{y} = \frac{52.72}{8} = 6.59 \qquad s_y = \sqrt{\frac{347.4814 - (52.72)^2/8}{8-1}} = 0.089921$$

El estadístico t apropiado se calcula como:

$$t_{0.05/2,8-1} = t_{0.025,7} = 2.364623$$

que se utiliza para calcular el intervalo

$$L = 6.59 - \frac{0.089921}{\sqrt{8}}2.364623 = 6.5148$$

$$U = 6.59 + \frac{0.089921}{\sqrt{8}}2.364623 = 6.6652$$

o

$$6.5148 \leq \mu \leq 6.6652$$

Así, considerando las primeras ocho mediciones, concluimos que existe un 95% de probabilidad de que la media real esté en el intervalo de 6.5148 a 6.6652.

Los otros dos casos, *b*) con 16 valores y *c*) con 24 valores, se calculan en forma similar y los resultados se tabulan junto con los del inciso *a*) como sigue:

n	$\bar{y}$	s_y	$t_{\alpha/2,n-1}$	L	U
8	6.5900	0.089921	2.364623	6.5148	6.6652
16	6.5794	0.095845	2.131451	6.5283	6.6304
24	6.6000	0.097133	2.068655	6.5590	6.6410

Estos resultados, que también se resumen en la figura PT5.5, indican el resultado esperado de que el intervalo de confianza se vuelve más pequeño conforme n aumenta. Así, cuantas más mediciones se tomen, nuestra estimación del valor verdadero será más refinado.

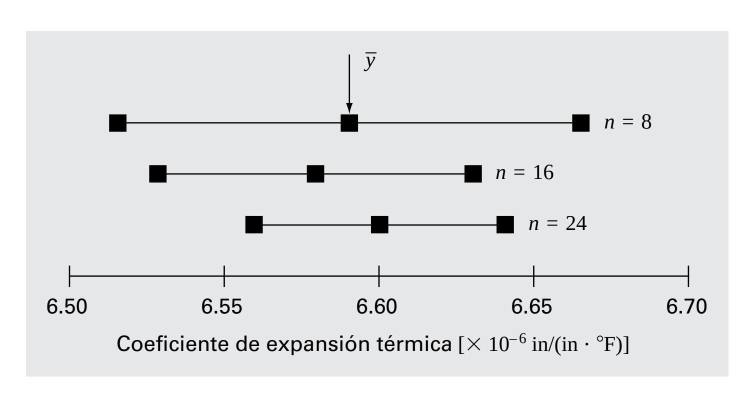


Figura PT5.5
Estimaciones de la media e intervalos de confianza del 95% para diferentes tamaños de la muestra.

Lo anterior es sólo un simple ejemplo de cómo se emplea la estadística para tomar decisiones respecto de datos inciertos. Esos conceptos también serán de relevancia en nuestro análisis de los modelos de regresión. Usted puede consultar cualquier libro básico de estadística (por ejemplo, Milton y Arnold, 2002) para obtener más información sobre este tema.

PT5.3 ORIENTACIÓN

Antes de proceder con los métodos numéricos para el ajuste de curvas, la siguiente orientación podría ser de utilidad. Este apartado se presenta como una visión general del material que se estudia en la parte cinco. Además, se formulan algunos objetivos para ayudar a enfocar su atención al analizar el tema.

PT5.3.1 Alcance y presentación preliminar

La figura PT5.6 proporciona una visión general del material que se estudiará en la parte cinco. El *capítulo 17* se dedica a la *regresión por mínimos cuadrados*. Se aprenderá primero cómo ajustar la "mejor" línea recta a través de un conjunto de datos inciertos. Esta técnica se conoce como *regresión lineal*. Además de analizar cómo calcular la pendiente y la intersección, con el eje y, de esta línea recta, se presentarán también métodos visuales y cuantitativos para evaluar la validez de los resultados.

Además de ajustar a una línea recta, se mostrará también una técnica general para ajustar a un "mejor" polinomio. Así, usted aprenderá a obtener un polinomio parabólico, cúbico o de un grado superior, que se ajuste en forma óptima a datos inciertos. La regresión lineal es un subconjunto de este procedimiento más general, llamado *regresión polinomial*.

El siguiente tema que se analiza en el capítulo 17 es la *regresión lineal múltiple*. Está diseñada para el caso donde la variable dependiente y es una función lineal de dos o más variables independientes $x_1, x_2, \ldots, x_m$. Este procedimiento tiene especial utilidad para evaluar datos experimentales donde la variable de interés es dependiente de varios factores.

Después de la regresión múltiple, ilustramos cómo tanto la regresión polinomial como la múltiple son subconjuntos de un *modelo lineal general de mínimos cuadrados*. Entre otras cuestiones, esto nos permitirá introducir una representación matricial concisa de la regresión y analizar sus propiedades estadísticas generales.

Por último, las últimas secciones del capítulo 17 se dedican a la *regresión no lineal*. Este procedimiento está diseñado para calcular un ajuste por mínimos cuadrados de una ecuación no lineal a datos.

En el *capítulo 18* se describe la técnica alternativa para el ajuste de curvas llamada *interpolación*. Como se analizó antes, la interpolación se utiliza para estimar valores intermedios entre datos precisos. En el capítulo 18 se obtienen polinomios con este propósito. Se introduce el concepto básico de interpolación polinomial usando líneas rectas y parábolas para unir los puntos. Después, se desarrolla un procedimiento generalizado para ajustar un polinomio de grado n. Se presentan dos métodos para expresar tales polinomios en forma de ecuación. El primero, llamado *interpolación polinomial de Newton*, es preferible cuando se desconoce el grado apropiado del polinomio. El segundo, llamado *interpolación polinomial de Lagrange*, tiene ventajas cuando de antemano se conoce el grado apropiado.

La siguiente sección del capítulo 18 presenta una técnica alternativa para ajustar datos precisos. Ésta, llamada *interpolación mediante trazadores o splines*, ajusta polinomios a datos, pero en forma de trozos. Como tal, es particularmente adecuada para ajustar datos que en general son suaves pero que muestren abruptos cambios locales. Finalmente, se da una breve introducción a la interpolación multidimensional.

El *capítulo 19* tiene que ver con el método de la transformada de Fourier para el ajuste de curvas, donde funciones periódicas se ajustan a datos. Nuestro énfasis en esta sección residirá en la *transformada rápida de Fourier.* Al final se incluye también una revisión de algunos paquetes de software que se utilizan para el ajuste de curvas; entre ellos se encuentran Excel, MATLAB y Mathcad.

El *capítulo 20* se dedica a aplicaciones en la ingeniería que ilustran la utilidad de los métodos numéricos en el contexto de los problemas de ingeniería. Los ejemplos se toman de las cuatro áreas principales de la ingeniería: química, civil, eléctrica y mecánica. Además, algunas de las aplicaciones ilustran cómo se emplean los paquetes de software para resolver problemas de ingeniería.

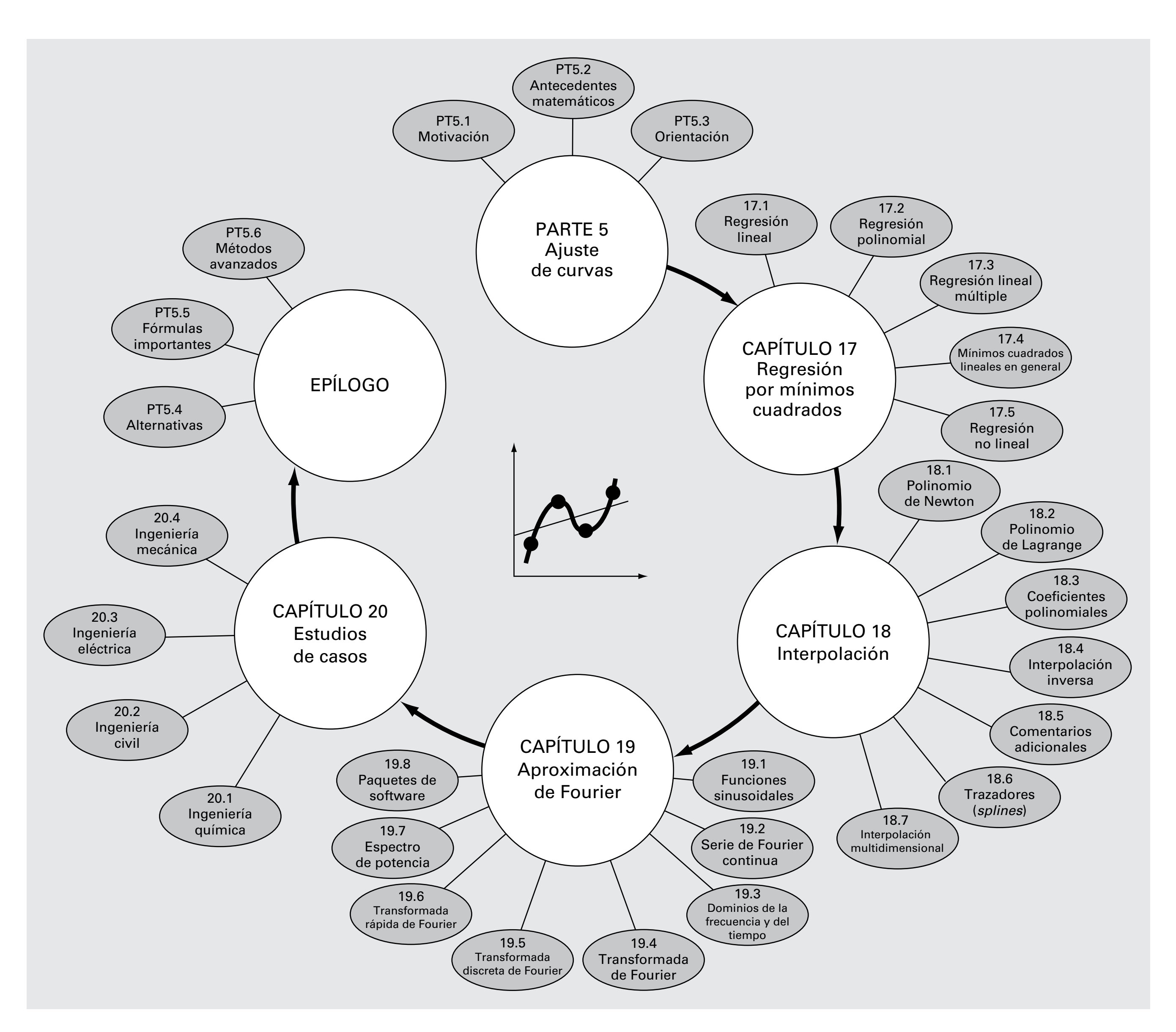


Figura PT5.6
Representación esquemática de la organización del material en la parte cinco: Ajuste de curvas.

Por último, se incluye un epílogo al final de la parte cinco. Contiene un resumen de las fórmulas y los conceptos importantes relacionados con el ajuste de curvas, así como un análisis de las ventajas y desventajas de las técnicas, y sugerencias para futuros estudios.

PT5.3.2 Metas y objetivos

Objetivos de estudio Después de estudiar la parte cinco, usted habrá mejorado su capacidad para ajustar curvas a los datos. En general, usted dominará las técnicas, habrá aprendido a valorar

la confiabilidad de los resultados y será capaz de seleccionar el método (o métodos) para cualquier problema específico. Además de estas metas generales, los conceptos particulares de la tabla PT5.3 deberán asimilarse y dominarse.

Objetivos computacionales Se le han proporcionado algoritmos de cómputo simples para implementar las técnicas analizadas en la parte cinco. También usted puede tener acceso a los paquetes de software. Todo esto tiene utilidad como herramientas de aprendizaje.

Se proporcionan algoritmos en pseudocódigo para la mayoría de los métodos en la parte cinco. Esta información le permitirá expandir sus bibliotecas de software para incluir técnicas más allá de la regresión polinomial. Por ejemplo, usted puede encontrar útil, desde un punto de vista profesional, tener software para la regresión lineal múltiple, la interpolación polinomial de Newton, la interpolación con trazadores cúbicos y la transformada rápida de Fourier.

Además, una de las metas más importantes deberá ser dominar varios de los paquetes de software de utilidad general que están disponibles. En particular, usted debería acostumbrarse a usar esas herramientas para implementar métodos numéricos en la solución de problemas en ingeniería.

Tabla PT5.3 Objetivos específicos de estudio de la parte cinco.

1. Comprender la diferencia fundamental entre regresión e interpolación, y darse cuenta de que confundirlos puede llevar a serios problemas.
2. Entender la deducción de la regresión lineal por mínimos cuadrados y ser capaz de evaluar la confiabilidad del ajuste mediante evaluaciones gráficas y cuantitativas.
3. Saber cómo linearizar datos mediante transformación.
4. Entender situaciones donde son apropiadas las regresiones polinomiales, múltiples y no lineales.
5. Ser capaz de reconocer modelos lineales generales, entender la formulación matricial general para mínimos cuadrados lineales, y saber cómo calcular intervalos de confianza para parámetros.
6. Entender que hay uno y sólo un polinomio de grado n o menor que pasa exactamente a través de $n + 1$ puntos.
7. Saber cómo obtener el polinomio de interpolación de Newton de primer grado.
8. Reconocer la analogía entre el polinomio de Newton y la expansión de la serie de Taylor, y cómo se relaciona el error de truncamiento.
9. Comprender que las ecuaciones de Newton y Lagrange son simplemente formulaciones diferentes de la misma interpolación polinomial, y entender sus respectivas ventajas y desventajas.
10. Percatarse de que, por lo general, se obtienen resultados más exactos si los datos usados para interpolación están más o menos centrados y cercanos al punto desconocido.
11. Darse cuenta que los datos no tienen que estar igualmente espaciados ni en un orden particular para los polinomios de Newton o de Lagrange.
12. Saber por qué son útiles las fórmulas de interpolación con igual espaciamiento.
13. Reconocer las desventajas y los riesgos asociados con la extrapolación.
14. Entender por qué los *trazadores* (*splines*) tienen utilidad para datos con áreas locales de cambio abrupto.
15. Entender cómo se puede aplicar en dos dimensiones la interpolación de polinomios.
16. Reconocer cómo se usa la serie de Fourier para ajustar datos a funciones periódicas.
17. Entender la diferencia entre dominios de la frecuencia y del tiempo.

CAPÍTULO 17

Regresión por mínimos cuadrados

Cuando los datos tienen errores sustanciales, la interpolación polinomial es inapropiada y puede dar resultados poco satisfactorios cuando se utiliza para predecir valores intermedios. Con frecuencia los datos experimentales son de este tipo. Por ejemplo, en la figura 17.1*a* se muestran siete datos obtenidos experimentalmente que presentan una variabilidad significativa. Una inspección visual de esos datos sugiere una posible relación entre y y x. Es decir, la tendencia general indica que valores altos de y están asociados con valores altos de x. Ahora, si un polinomio de interpolación de sexto grado se ajusta a estos datos (figura 17.1*b*), pasará exactamente a través de todos los puntos. Sin embargo, a causa de la variabilidad en los datos, la curva oscila mucho en el intervalo entre los puntos. En particular, los valores interpolados para $x = 1.5$ y $x = 6.5$ parecen estar bastante más allá del rango sugerido por los datos.

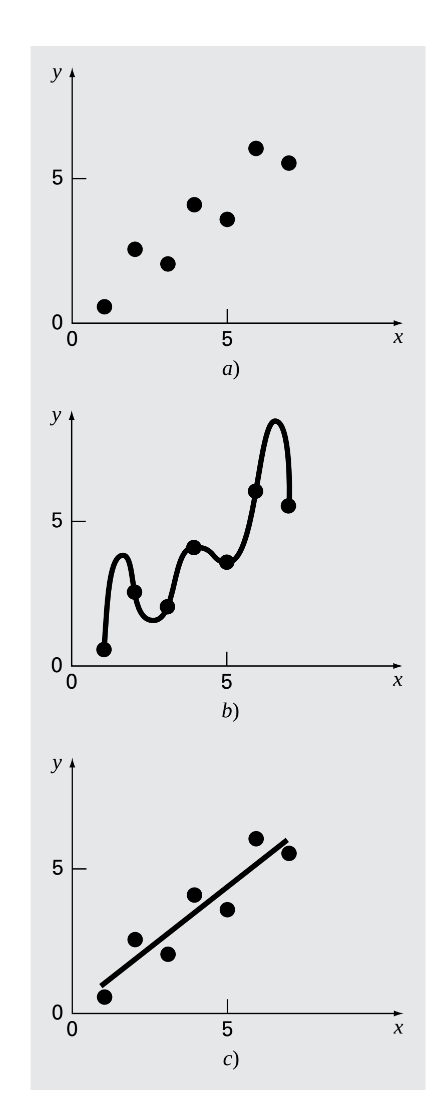


Figura 17.1
a) Datos que muestran un error significativo. *b*) Ajuste polinomial oscilando más allá del rango de los datos. *c*) Resultados más satisfactorios mediante el ajuste por mínimos cuadrados.

Una estrategia más apropiada en tales casos consiste en obtener una función de aproximación que se ajuste a la forma o a la tendencia general de los datos, sin coincidir necesariamente en todos los puntos. La figura 17.1*c* ilustra cómo se utiliza una línea recta para caracterizar de manera general la tendencia de los datos sin pasar a través de algún punto específico.

Una manera para determinar la línea de la figura 17.1*c* es inspeccionar en forma visual los datos graficados y después trazar una “mejor” línea a través de los puntos. Aunque tales procedimientos “a ojo” apelan al sentido común y son válidos para cálculos “superficiales”, resultan deficientes por ser arbitrarios. Es decir, a menos que los puntos definan una línea recta perfecta (en cuyo caso la interpolación resultaría apropiada), diferentes analistas dibujarían líneas distintas.

Para dejar a un lado dicha subjetividad se debe encontrar algún criterio para establecer una base para el ajuste. Una forma de hacerlo es obtener una curva que minimice la discrepancia entre los puntos y la curva. Una técnica para lograr tal objetivo, llamada *regresión por mínimos cuadrados*, se analizará en este capítulo.

17.1 REGRESIÓN LINEAL



El ejemplo más simple de una aproximación por mínimos cuadrados es ajustar una línea recta a un conjunto de observaciones definidas por puntos: (x_1, y_1), (x_2, y_2),…, (x_n, y_n). La expresión matemática para la línea recta es

$$y = a_0 + a_1x + e \tag{17.1}$$

donde a_0 y a_1 son coeficientes que representan la intersección con el eje y y la pendiente, respectivamente, e es el error, o diferencia, entre el modelo y las observaciones, el cual se representa al reordenar la ecuación (17.1) como

$$e = y - a_0 - a_1x$$

Así, el *error* o *residuo* es la discrepancia entre el valor verdadero de y y el valor aproximado, $a_0 + a_1x$, que predijo la ecuación lineal.

17.1.1 Criterio para un “mejor” ajuste

Una estrategia para ajustar una “mejor” línea a través de los datos será minimizar la suma de los errores residuales de todos los datos disponibles, como sigue:

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i) \tag{17.2}$$

donde n = número total de puntos. Sin embargo, éste es un criterio inadecuado, como lo muestra la figura 17.2*a*, la cual presenta el ajuste de una línea recta de dos puntos. Obviamente, el mejor ajuste es la línea que une los puntos. Sin embargo, cualquier línea recta que pase a través del punto medio que une la línea (excepto una línea perfectamente vertical) da como resultado un valor mínimo de la ecuación (17.2) igual a cero, debido a que los errores se cancelan.

Por lo tanto, otro criterio lógico podría ser minimizar la suma de los valores absolutos de las diferencias,

$$\sum_{i=1}^{n} |e_i| = \sum_{i=1}^{n} |y_i - a_0 - a_1 x_i|$$

La figura 17.2*b* muestra por qué este criterio también es inadecuado. Para los cuatro puntos dados, cualquier línea recta que esté dentro de las líneas punteadas minimizará el valor absoluto de la suma. Así, este criterio tampoco dará un único mejor ajuste.

Una tercera estrategia para ajustar una mejor línea es el criterio *minimax*. En esta técnica, la línea se elige de manera que minimice la máxima distancia a que un punto se encuentra de la línea. Como se ilustra en la figura 17.2*c*, tal estrategia es inadecuada para la regresión, ya que da excesiva influencia a puntos fuera del conjunto; es decir, a un solo punto con un gran error. Deberá observarse que el principio minimax es, en algunas ocasiones, adecuado para ajustar una función simple a una función complicada (Carnahan, Luther y Wilkes, 1969).

La estrategia que supera las deficiencias de los procedimientos mencionados consiste en minimizar la suma de los cuadrados de los residuos entre la y medida y la y calculada con el modelo lineal

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_{i,\text{medida}} - y_{i,\text{modelo}})^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 \tag{17.3}$$

Este criterio tiene varias ventajas, entre ellas el hecho de que se obtiene una línea única para cierto conjunto de datos. Antes de analizar tales propiedades, presentaremos una técnica para determinar los valores de a_0 y a_1 que minimizan la ecuación (17.3).

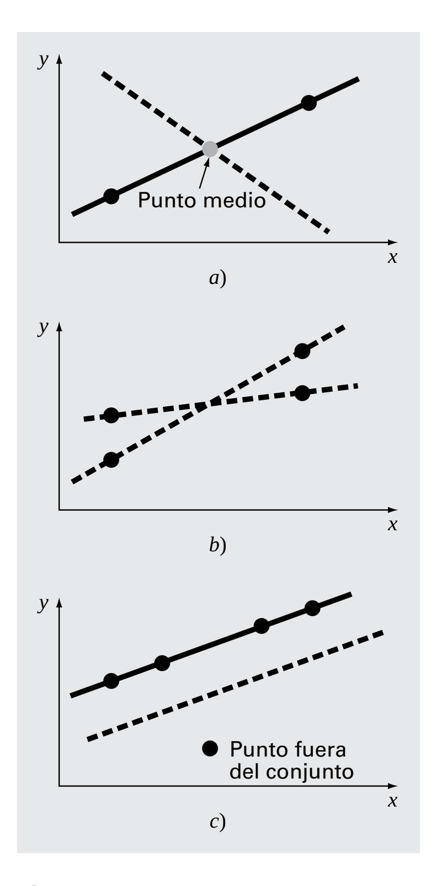


Figura 17.2
Ejemplo de algunos criterios para "el mejor ajuste" que son inadecuados para la regresión: *a*) minimizar la suma de los residuos, *b*) minimizar la suma de los valores absolutos de los residuos y *c*) minimizar el error máximo de cualquier punto individual.

17.1.2 Ajuste de una línea recta por mínimos cuadrados

Para determinar los valores de a_0 y a_1, la ecuación (17.3) se deriva con respecto a cada uno de los coeficientes:

$$\frac{\partial S_r}{\partial a_0} = -2\sum (y_i - a_0 - a_1 x_i)$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum [(y_i - a_0 - a_1 x_i)x_i]$$

Observe que hemos simplificado los símbolos de la sumatoria; a menos que se indique otra cosa, todas las sumatorias van desde $i = 1$ hasta n. Al igualar estas derivadas a cero, se dará como resultado un S_r mínimo. Si se hace esto, las ecuaciones se expresan como

$$0 = \sum y_i - \sum a_0 - \sum a_1 x_i$$

$$0 = \sum y_i x_i - \sum a_0 x_i - \sum a_1 x_i^2$$

Ahora, si observamos que $\sum a_0 = na_0$, expresamos las ecuaciones como un conjunto de dos ecuaciones lineales simultáneas, con dos incógnitas (a_0 y a_1):

$$na_0 + \left(\sum x_i\right) a_1 = \sum y_i \tag{17.4}$$

$$\left(\sum x_i\right) a_0 + \left(\sum x_i^2\right) a_1 = \sum x_i y_i \tag{17.5}$$

Éstas se llaman *ecuaciones normales*, y se resuelven en forma simultánea

$$a_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \tag{17.6}$$

Este resultado se utiliza conjuntamente con la ecuación (17.4) para obtener

$$a_0 = \bar{y} - a_1 \bar{x} \tag{17.7}$$

donde $\bar{y}$ y $\bar{x}$ son las medias de y y x, respectivamente.

EJEMPLO 17.1 Regresión lineal

Planteamiento del problema Ajuste a una línea recta los valores x y y en las dos primeras columnas de la tabla 17.1.

Tabla 17.1 Cálculos para el análisis de error en el ajuste lineal.

x_i	y_i	$(y_i - \bar{y})$	$(y_i - a_0 - a_1 x_i)^2$
1	0.5	8.5765	0.1687
2	2.5	0.8622	0.5625
3	2.0	2.0408	0.3473
4	4.0	0.3265	0.3265
5	3.5	0.0051	0.5896
6	6.0	6.6122	0.7972
7	5.5	4.2908	0.1993
Σ	24.0	22.7143	2.9911

Solución Se calculan las siguientes cantidades:

$$n = 7 \qquad \sum x_i y_i = 119.5 \qquad \sum x_i^2 = 140$$

$$\sum x_i = 28 \qquad \bar{x} = \frac{28}{7} = 4$$

$$\sum y_i = 24 \qquad \bar{y} = \frac{24}{7} = 3.428571$$

Mediante las ecuaciones (17.6) y (17.7)

$$a_1 = \frac{7(119.5) - 28(24)}{7(140) - (28)^2} = 0.8392857$$

$$a_0 = 3.428571 - 0.8392857(4) = 0.07142857$$

Por lo tanto, el ajuste por mínimos cuadrados es

$$y = 0.07142857 + 0.8392857x$$

La línea, junto con los datos, se muestran en la figura 17.1*c*.

17.1.3 Cuantificación del error en la regresión lineal

Cualquier otra línea diferente a la calculada en el ejemplo 17.1 dará como resultado una suma mayor de los cuadrados de los residuos. Así, la línea es única y, en términos de nuestro criterio elegido, es la "mejor" línea a través de los puntos. Varias propiedades de este ajuste se observan al examinar más de cerca la forma en que se calcularon los residuos. Recuerde que la suma de los cuadrados se define como [ecuación (17.3)]

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 \tag{17.8}$$

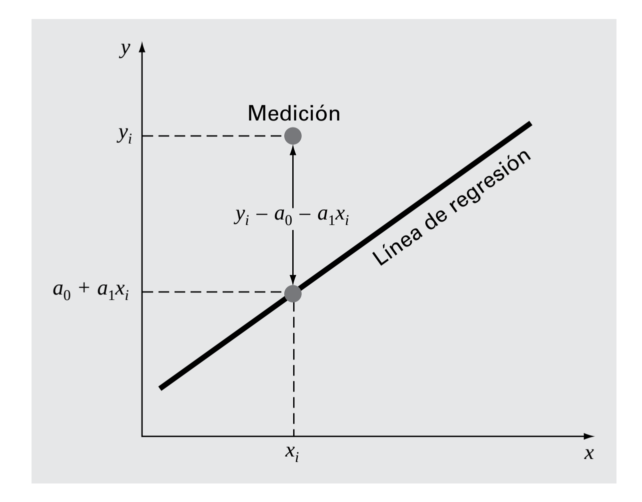


Figura 17.3
El residuo en la regresión lineal representa la distancia vertical entre un dato y la línea recta.

Observe la similitud entre las ecuaciones (PT5.3) y (17.8). En el primer caso, el cuadrado del residuo representa el cuadrado de la diferencia entre el dato y una estimación de la medida de tendencia central: la media. En la ecuación (17.8), el cuadrado del residuo representa el cuadrado de la distancia vertical entre el dato y otra medida de tendencia central: la línea recta (figura 17.3).

La analogía se puede extender aún más en casos donde 1) la dispersión de los puntos alrededor de la línea es de magnitud similar en todo el rango de los datos, y 2) la distribución de estos puntos cerca de la línea es normal. Es posible demostrar que si estos criterios se cumplen, la regresión por mínimos cuadrados proporcionará la mejor (es decir, la más adecuada) estimación de a_0 y a_1 (Draper y Smith, 1981). Esto se conoce en estadística como el *principio de máxima verosimilitud*. Además, si estos criterios se satisfacen, una "desviación estándar" para la línea de regresión se determina como sigue [compare con la ecuación (PT5.2)]

$$s_{y/x} = \sqrt{\frac{S_r}{n-2}} \tag{17.9}$$

donde a $s_{y/x}$ se le llama *error estándar de la estimación*. El subíndice "y/x" designa que el error es para un valor predicho de y correspondiente a un valor particular de x. También, observe que ahora dividimos entre $n - 2$ debido a que se usaron dos datos estimados (a_0 y a_1), para calcular S_r; así, se han perdido dos grados de libertad. Como lo hicimos en nuestro análisis para la desviación estándar en PT5.2.1, otra justificación para dividir entre $n - 2$ es que no existe algo como "datos dispersos" alrededor de una línea recta que une dos puntos. De esta manera, en el caso donde $n = 2$, la ecuación (17.9) da un resultado sin sentido, infinito.

Así como en el caso de la desviación estándar, el error estándar de la estimación cuantifica la dispersión de los datos. Aunque, $s_{y/x}$ cuantifica la dispersión *alrededor de la línea de regresión*, como se muestra en la figura 17.4*b*, a diferencia de la desviación estándar original s_y que cuantifica la dispersión *alrededor de la media* (figura 17.4*a*).

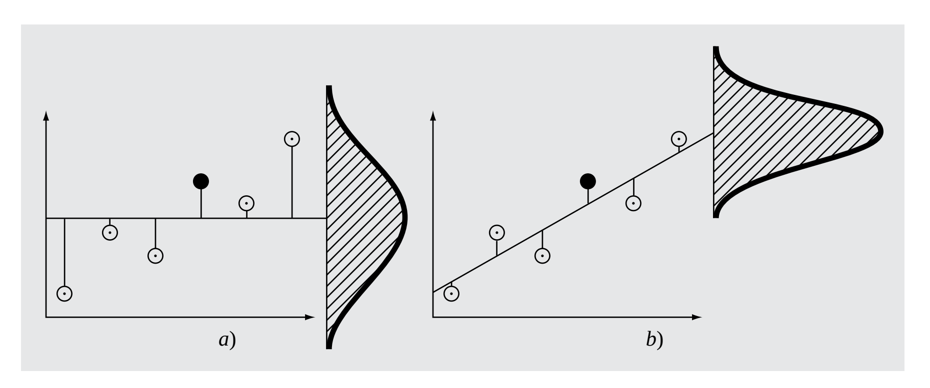


Figura 17.4
Datos de regresión que muestran *a*) la dispersión de los datos alrededor de la media de la variable dependiente y *b*) la dispersión de los datos alrededor de la línea de mejor ajuste. La reducción en la dispersión al ir de *a*) a *b*), como lo indican las curvas en forma de campana a la derecha, representa la mejora debida a la regresión lineal.

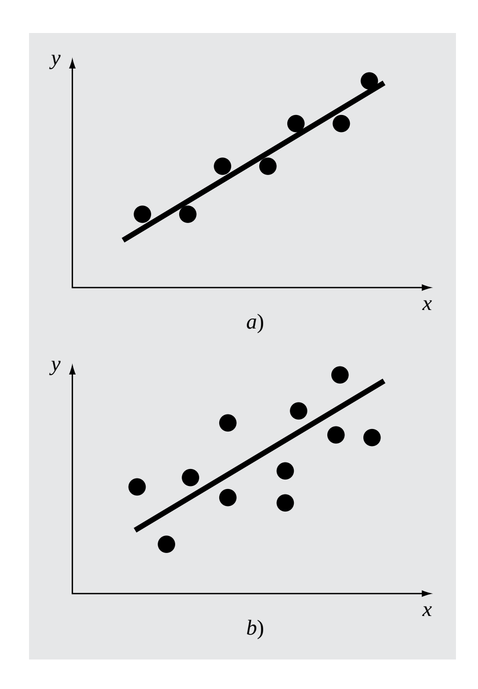


Figura 17.5
Ejemplos de regresión lineal con errores residuales *a*) pequeños y *b*) grandes.

Los conceptos anteriores se utilizan para cuantificar la "bondad" de nuestro ajuste. Esto es en particular útil para comparar diferentes regresiones (figura 17.5). Para hacerlo, regresamos a los datos originales y determinamos la *suma total de los cuadrados* alrededor de la media para la variable dependiente (en nuestro caso, *y*). Como en el caso de la ecuación (PT5.3), esta cantidad se designa por S_t. Ésta es la magnitud del error residual asociado con la variable dependiente antes de la regresión. Después de realizar la regresión, calculamos S_r, es decir, la suma de los cuadrados de los residuos alrededor de la línea de regresión. Esto caracteriza el error residual que queda después de la regresión. Es por lo que, algunas veces, se le llama la suma inexplicable de los cuadrados. La diferencia entre estas dos cantidades, $S_t - S_r$, cuantifica la mejora o reducción del error por describir los datos en términos de una línea recta en vez de un valor promedio. Como la magnitud de esta cantidad depende de la escala, la diferencia se normaliza a S_t para obtener

$$r^2 = \frac{S_t - S_r}{S_t} \tag{17.10}$$

donde r^2 se conoce como el *coeficiente de determinación* y *r* es el *coeficiente de correlación* ($= \sqrt{r^2}$). En un ajuste perfecto, $S_r = 0$ y $r = r^2 = 1$, significa que la línea explica el 100% de la variabilidad de los datos. Si $r = r^2 = 0$, $S_r = S_t$ el ajuste no representa alguna mejora. Una representación alternativa para *r* que es más conveniente para implementarse en una computadora es

$$r = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}} \tag{17.11}$$

EJEMPLO 17.2 Estimación de errores en el ajuste lineal por mínimos cuadrados

Planteamiento del problema Calcule la desviación estándar total, el error estándar del estimado y el coeficiente de correlación para los datos del ejemplo 17.1.

Solución Las sumatorias se realizan y se presentan en la tabla 17.1. La desviación estándar es [ecuación (PT5.2)]

$$s_y = \sqrt{\frac{22.7143}{7-1}} = 1.9457$$

y el error estándar del estimado es [ecuación (17.9)]

$$s_{y/x} = \sqrt{\frac{2.9911}{7-2}} = 0.7735$$

Como $s_{y/x} < s_y$, el modelo de regresión lineal es adecuado. La mejora se puede cuantificar mediante [ecuación (17.10)]

$$r^2 = \frac{22.7143 - 2.9911}{22.7143} = 0.868$$

o

$$r = \sqrt{0.868} = 0.932$$

Los resultados indican que el modelo lineal explicó el 86.8% de la incertidumbre original.

Antes de implementar el programa computacional para la regresión lineal, debemos tomar en cuenta algunas consideraciones. Aunque el coeficiente de correlación ofrece una manera fácil de medir la bondad del ajuste, se deberá tener cuidado de no darle más significado del que ya tiene. El solo hecho de que r sea "cercana" a 1 no necesariamente significa que el ajuste sea "bueno". Por ejemplo, es posible obtener un valor relativamente alto de r cuando la relación entre y y x no es lineal. Draper y Smith (1981) proporcionan guías y material adicional respecto a la evaluación de resultados en la regresión lineal. Además, como mínimo, usted deberá inspeccionar *siempre* una gráfica de los datos junto con su curva de regresión. Como se describe en la siguiente sección, los paquetes de software tienen estas capacidades.

17.1.4 Programa computacional para la regresión lineal

Es relativamente fácil desarrollar un pseudocódigo para la regresión lineal (figura 17.6). Como se mencionó antes, la opción de graficar resulta benéfico para el uso efectivo y la interpretación de la regresión. Tales capacidades se incluyen en paquetes de software populares como Excel y MATLAB. Si su lenguaje de computación tiene capacidad para graficar, recomendamos que expanda su programa para incluir una gráfica de y contra x, que muestre tanto los datos como la línea de regresión. La inclusión de la capacidad aumentará mucho la utilidad del programa en los contextos de solución de problemas.

EJEMPLO 17.3 Regresión lineal usando la computadora

Planteamiento del problema Se utiliza el software basado en la figura 17.6 para resolver un problema de prueba de hipótesis relacionado con la caída del paracaidista que se analizó en el capítulo 1. Un modelo teórico matemático para la velocidad del paracaidista se dio como sigue [ecuación (1.10)]:

$$v(t) = \frac{gm}{c}\left(1 - e^{(-c/m)t}\right)$$

```
SUB Regress(x, y, n, a1, a0, syx, r2)

  sumx = 0: sumxy = 0: st = 0
  sumy = 0: sumx2 = 0: sr = 0
  DOFOR i = 1, n
    sumx = sumx + x_i
    sumy = sumy + y_i
    sumxy = sumxy + x_i*y_i
    sumx2 = sumx2 + x_i*x_i
  END DO
  xm = sumx/n
  ym = sumy/n
  a1 = (n*sumxy − sumx*sumy)/(n*sumx2 − sumx*sumx)
  a0 = ym − a1*xm
  DOFOR i = 1, n
    st = st + (y_i − ym)^2
    sr = sr + (y_i − a1*x_i − a0)^2
  END DO
  syx = (sr/(n − 2))^0.5
  r2 = (st − sr)/st

END Regress
```

Figura 17.6
Algoritmo para la regresión lineal.

donde v = velocidad (m/s), g = constante gravitacional (9.8 m/s^2), m = masa del paracaidista igual a 68.1 kg y c = coeficiente de arrastre de 12.5 kg/s. El modelo predice la velocidad del paracaidista en función del tiempo, como se describe en el ejemplo 1.1.

Un modelo empírico alternativo para la velocidad del paracaidista está dado por

$$v(t) = \frac{gm}{c}\left(\frac{t}{3.75+t}\right) \tag{E17.3.1}$$

Suponga que usted quiere probar y comparar la veracidad de esos dos modelos matemáticos. Esto se podría hacer al medir la velocidad real del paracaidista con valores conocidos de tiempo y al comparar estos resultados con las velocidades predichas de acuerdo con cada modelo.

Se implementó un programa para la recolección de datos experimentales, y los resultados se enlistan en la columna *a*) de la tabla 17.2. Las velocidades calculadas con cada modelo se enlistan en las columnas *b*) y *c*).

Solución La veracidad de los modelos se prueba al graficar la velocidad calculada por el modelo contra la velocidad medida. Se puede usar la regresión lineal para calcular la pendiente y la intersección con el eje y de la gráfica. Esta línea tendrá una pendiente de 1, una intersección de 0 y $r^2 = 1$ si el modelo concuerda perfectamente con los datos. Una desviación significativa de estos valores sirve como una indicación de lo inadecuado del modelo.

Las figuras 17.7*a* y *b* muestran gráficas de la línea y los datos para las regresiones de las columnas *b*) y *c*), respectivamente, contra la columna *a*). Para el primer modelo [ecuación (1.10) como se ilustra en la figura 17.7*a*]

Tabla 17.2 Velocidades medidas y calculadas para la caída del paracaidista.

Tiempo, s	*v* medida, m/s *a*)	*v* calculada con el modelo, m/s [ec. (1.10)] *b*)	*v* calculada con el modelo, m/s [ec. (E17.3.1)] *c*)
1	10.00	8.953	11.240
2	16.30	16.405	18.570
3	23.00	22.607	23.729
4	27.50	27.769	27.556
5	31.00	32.065	30.509
6	35.60	35.641	32.855
7	39.00	38.617	34.766
8	41.50	41.095	36.351
9	42.90	43.156	37.687
10	45.00	44.872	38.829
11	46.00	46.301	39.816
12	45.50	47.490	40.678
13	46.00	48.479	41.437
14	49.00	49.303	42.110
15	50.00	49.988	42.712

$$v_{\text{modelo}} = -0.859 + 1.032 v_{\text{medida}}$$

y para el segundo modelo [ecuación (E17.3.1) como se ilustra en la figura 17.7*b*],

$$v_{\text{modelo}} = 5.776 + 0.752 v_{\text{medida}}$$

Esas gráficas indican que la regresión lineal entre los datos y cada uno de los modelos es altamente significativa. Ambos modelos ajustan los datos con un coeficiente de correlación mayor a 0.99.

No obstante, el modelo descrito por la ecuación (1.10) se ajusta mejor a nuestro criterio de prueba de hipótesis que el descrito por la ecuación (E17.3.1), ya que la pendiente y la intersección con el

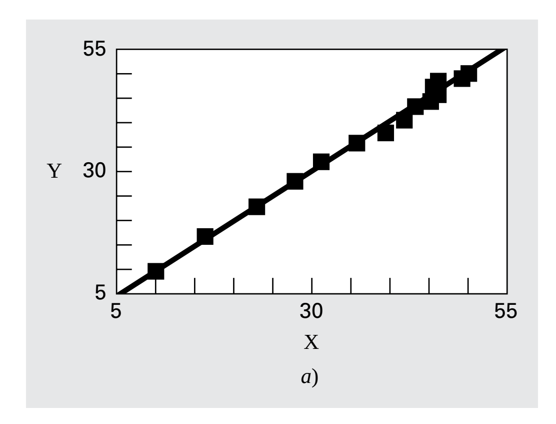
eje *y* son más cercanos a 1 y 0. Así, aunque cada gráfica queda bien descrita por una línea recta, la ecuación (1.10) parece ser un mejor modelo que la (E17.3.1).

La prueba y la selección del modelo son actividades comunes y muy importantes en todas las ramas de la ingeniería. El material que se presentó antes en este capítulo, junto con su software, le ayudarán a resolver muchos problemas prácticos de este tipo.

El análisis en el ejemplo 17.3 tiene un defecto: el ejemplo no fue ambiguo, ya que el modelo empírico [ecuación (E17.3.1)] fue claramente inferior al de la ecuación (1.10). La pendiente y la intersección en el modelo empírico fueron mucho más cercanos a los resultados deseados 1 y 0, por lo que resultó obvio cuál era el mejor modelo.

Sin embargo, suponga que la pendiente fuera de 0.85 y que la intersección con el eje *y* fuera de 2. Obviamente esto llevaría a la conclusión de que la pendiente y la intersección fueran 1 y 0 respectivamente. Por lo anterior, es claro que, más que apoyarse en un juicio subjetivo, es preferible basar tal conclusión sobre un criterio cuantitativo.

Esto se logra al calcular intervalos de confianza para los parámetros del modelo, de la misma forma que desarrollamos intervalos de confianza para la media en la sección PT5.2.3. Regresaremos a este punto al final del capítulo.



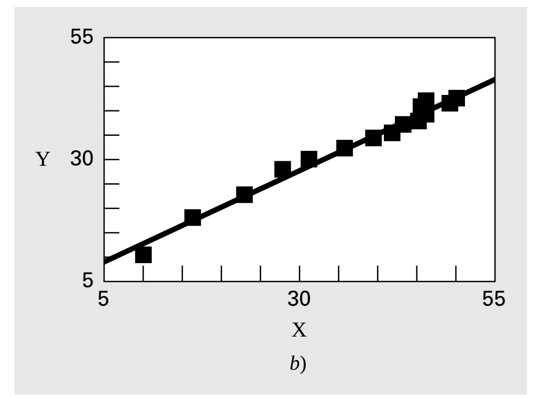


Figura 17.7
a) Resultados con regresión lineal para comparar las predicciones calculadas con el modelo teórico [ecuación (1.10)] contra valores medidos. *b*) Resultados con regresión lineal para comparar predicciones calculadas con el modelo empírico [ecuación (E17.3.1] contra valores medidos.

17.1.5 Linealización de relaciones no lineales

La regresión lineal ofrece una poderosa técnica para ajustar una mejor línea a los datos. Sin embargo, se considera el hecho de que la relación entre las variables dependiente e independiente es lineal. Éste no es siempre el caso, y el primer paso en cualquier análisis de regresión deberá ser graficar e inspeccionar los datos en forma visual, para asegurarnos que sea posible usar un modelo lineal. Por ejemplo, la figura 17.8 muestra algunos datos que obviamente son curvilíneos. En algunos casos, las técnicas como la regresión polinomial, que se describen en la sección 17.2, son apropiadas. En otros, se pueden utilizar transformaciones para expresar los datos en una forma que sea compatible con la regresión lineal.

Un ejemplo es el *modelo exponencial*

$$y = \alpha_1 e^{\beta_1 x} \tag{17.12}$$

donde α_1 y β_1 son constantes. Este modelo se emplea en muchos campos de la ingeniería para caracterizar cantidades que aumentan (β_1 positivo) o disminuyen (β_1 negativo), a una velocidad que es

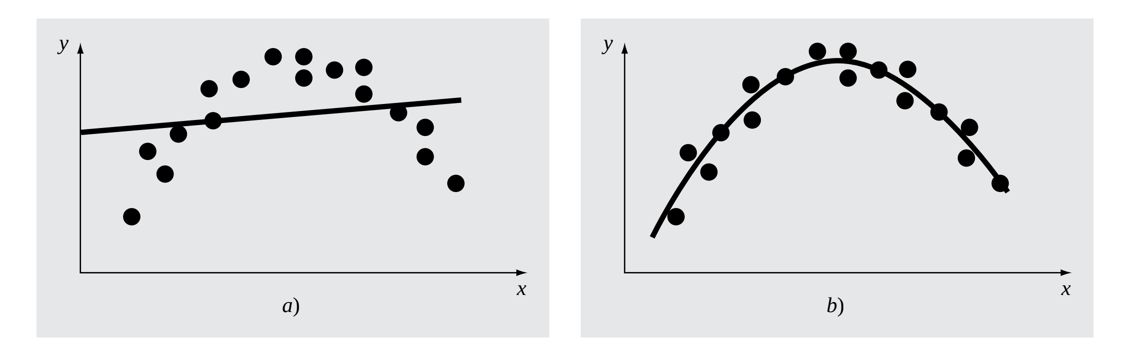


Figura 17.8
a) Datos inadecuados para la regresión lineal por mínimos cuadrados. *b*) Indicación de que es preferible una parábola.

directamente proporcional a sus propias magnitudes. Por ejemplo, el crecimiento poblacional o el decaimiento radiactivo tienen este comportamiento. Como se ilustra en la figura 17.9*a*, la ecuación representa una relación no lineal (para $\beta_1 \neq 0$) entre *y* y *x*.

Figura 17.9
a) La ecuación exponencial, *b*) la ecuación de potencias y *c*) la ecuación de razón del crecimiento. Los incisos *d*), *e*) y *f*) son versiones linealizadas de estas ecuaciones que resultan de transformaciones simples.

Otro ejemplo de modelo no lineal es la ecuación de potencias

$$y = \alpha_2 x^{\beta_2} \tag{17.13}$$

donde α_2 y β_2 son coeficientes constantes. Este modelo tiene muchas aplicaciones en todos los campos de la ingeniería. Como se ilustra en la figura 17.9*b*, la ecuación (para $\beta_2 \neq 0$ o 1) es no lineal.

Un tercer ejemplo de un modelo no lineal es la ecuación de razón del crecimiento [recuerde la ecuación (E17.3.1)]

$$y = \alpha_3 \frac{x}{\beta_3 + x} \tag{17.14}$$

donde α_3 y β_3 son coeficientes constantes. Este modelo particularmente es adecuado para caracterizar la razón de crecimiento poblacional bajo condiciones limitantes, también representa una relación no lineal entre *y* y *x* (figura 17.9*c*) que se iguala o "satura", conforme *x* aumenta.

Hay técnicas de regresión no lineal disponibles para ajustar estas ecuaciones de manera directa a datos experimentales. (Observe que analizaremos la regresión no lineal en la sección 17.5.) Sin embargo, una alternativa simple consiste en usar manipulaciones matemáticas para transformar las ecuaciones en una forma lineal. Después, se utiliza la regresión lineal simple para ajustar las ecuaciones a los datos.

Por ejemplo, la ecuación (17.12) se linealiza al aplicarle el logaritmo natural obteniendo

$$\ln y = \ln \alpha_1 + \beta_1 x \ln e$$

Pero como ln *e* = l,

$$\ln y = \ln \alpha_1 + \beta_1 x \tag{17.15}$$

Así, una gráfica de ln y contra *x* dará una línea recta con una pendiente β_1 y una intersección con el eje de las ordenadas igual a ln α_1 (figura 17.9*d*).

La ecuación (17.13) es linealizada al aplicarle el logaritmo de base 10 obteniendo

$$\log y = \beta_2 \log x + \log \alpha_2 \tag{7.16}$$

De este modo, una gráfica de log *y* contra log *x* dará una línea recta con pendiente β_2 e intersección con el eje de las ordenadas log α_2 (figura 17.9*e*).

La ecuación (17.14) es linealizada al invertirla para dar

$$\frac{1}{y} = \frac{\beta_3}{\alpha_3}\frac{1}{x} + \frac{1}{\alpha_3} \tag{17.17}$$

De esta forma, una gráfica de 1/*y* contra 1/*x* será lineal, con pendiente β_3/α_3 y una intersección con el eje de las ordenadas $1/\alpha_3$ (figura 17.9*f*).

En sus formas transformadas, estos modelos pueden usar la regresión lineal para poder evaluar los coeficientes constantes. Después, regresarse a su estado original y usarse para fines predictivos. El ejemplo 17.4 ilustra este procedimiento con la ecuación (17.13). Además, la sección 20.1 proporciona un ejemplo de ingeniería de la misma clase de cálculo.

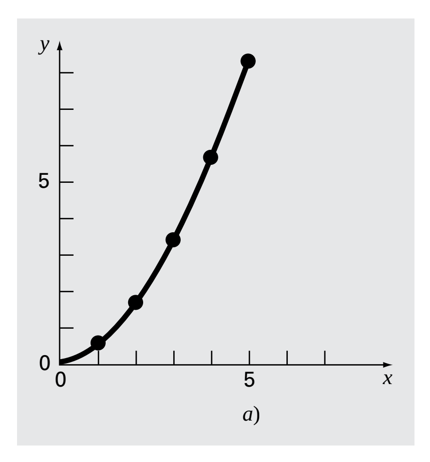
EJEMPLO 17.4 Linealización de una ecuación de potencias

Planteamiento del problema Ajuste la ecuación (17.13) a los datos de la tabla 17.3 mediante una transformación logarítmica de los datos.

Solución La figura l7.10*a* es una gráfica de los datos originales en su estado no transformado. La figura 17.10*b* muestra la gráfica de los datos transformados. Una regresión lineal de esta transformación mediante logaritmos dan el siguiente resultado:

Tabla 17.3 Datos que se ajustarán con la ecuación de potencias.

x	*y*	log *x*	log *y*
1	0.5	0	–0.301
2	1.7	0.301	0.226
3	3.4	0.477	0.534
4	5.7	0.602	0.753
5	8.4	0.699	0.922



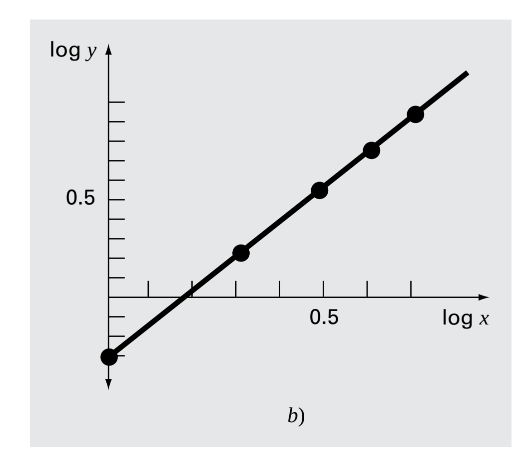


Figura 17.10
a) Gráfica de datos no transformados con la ecuación de potencias que se ajusta a los datos. *b*) Gráfica de datos transformados para determinar los coeficientes de la ecuación de potencias.

$$\log y = 1.75 \log x - 0.300$$

Así, la intersección con el eje de las ordenadas es log α_2 igual a –0.300 y, por lo tanto, al tomar el antilogaritmo, $\alpha_2 = 10^{-0.3} = 0.5$. La pendiente es $\beta_2 = 1.75$. En consecuencia, la ecuación de potencias es

$$y = 0.5x^{1.75}$$

Esta curva, como se grafica en la figura 17.10*a*, indica un buen ajuste.

17.1.6 Comentarios generales sobre la regresión lineal

Antes de plantear la regresión curvilínea y lineal múltiple, debemos enfatizar la naturaleza introductoria del material anterior sobre regresión lineal. Nos hemos concentrado en la obtención y el uso práctico de ecuaciones para ajustarse a datos. Deberá estar consciente del hecho de que hay aspectos teóricos de regresión que son de importancia práctica, pero que van más allá del alcance de este libro. Por ejemplo, algunas suposiciones estadísticas, inherentes a los procedimientos lineales por mínimos cuadrados, son

1. Cada *x* tiene un valor fijo; no es aleatorio y se conoce sin error.
2. Los valores de *y* son variables aleatorias independientes y todas tienen la misma varianza.
3. Los valores de *y* para una *x* dada deben estar distribuidos normalmente.

Tales suposiciones son relevantes para la obtención adecuada y el uso de la regresión. Por ejemplo, la primera suposición significa que 1) los valores *x* deben estar libres de errores, y 2) la regresión de *y* contra *x* no es la misma que la de *x* contra *y* (véase el problema 17.4 al final del capítulo). Usted

debe consultar otras referencias como Draper y Smith (1981) para apreciar los aspectos y detalles de la regresión que están más allá del alcance de este libro.

17.2 REGRESIÓN POLINOMIAL

En la sección 17.1 se desarrolló un procedimiento para obtener la ecuación de una línea recta por medio del criterio de mínimos cuadrados. En la ingeniería, aunque algunos datos exhiben un patrón marcado, como el que se advierte en la figura 17.8, son pobremente representados por una línea recta, entonces, una curva podrá ser más adecuada para ajustarse a los datos. Como se analizó en la sección anterior, un método para lograr este objetivo es utilizar transformaciones. Otra alternativa es ajustar polinomios a los datos mediante *regresión polinomial*.

El procedimiento de mínimos cuadrados se puede extender fácilmente al ajuste de datos con un polinomio de grado superior. Por ejemplo, suponga que ajustamos un polinomio de segundo grado o cuadrático:

$$y = a_0 + a_1x + a_2x^2 + e$$

En este caso, la suma de los cuadrados de los residuos es [compare con la ecuación (17.3)]

$$S_r = \sum_{i=1}^{n} (y_i - a_0 - a_1x_i - a_2x_i^2)^2 \tag{17.18}$$

Al seguir el procedimiento de la sección anterior, obtenemos la derivada de la ecuación (17.18) con respecto a cada uno de los coeficientes desconocidos del polinomio,

$$\frac{\partial S_r}{\partial a_0} = -2\sum (y_i - a_0 - a_1x_i - a_2x_i^2)$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum x_i(y_i - a_0 - a_1x_i - a_2x_i^2)$$

$$\frac{\partial S_r}{\partial a_2} = -2\sum x_i^2(y_i - a_0 - a_1x_i - a_2x_i^2)$$

Estas ecuaciones se igualan a cero y se reordenan para desarrollar el siguiente conjunto de ecuaciones normales:

$$\begin{aligned} (n)a_0 + \left(\sum x_i\right)a_1 + \left(\sum x_i^2\right)a_2 &= \sum y_i \\ \left(\sum x_i\right)a_0 + \left(\sum x_i^2\right)a_1 + \left(\sum x_i^3\right)a_2 &= \sum x_iy_i \\ \left(\sum x_i^2\right)a_0 + \left(\sum x_i^3\right)a_1 + \left(\sum x_i^4\right)a_2 &= \sum x_i^2y_i \end{aligned} \tag{17.19}$$

donde todas las sumatorias van desde $i = 1$ hasta n. Observe que las tres ecuaciones anteriores son lineales y tienen tres incógnitas: a_0, a_1 y a_2. Los coeficientes de las incógnitas se evalúan de manera directa, a partir de los datos observados.

En este caso, observamos que el problema de determinar un polinomio de segundo grado por mínimos cuadrados es equivalente a resolver un sistema de tres ecuaciones lineales simultáneas. En la parte tres se estudiaron las técnicas para resolver tales ecuaciones.

El caso bidimensional se extiende con facilidad a un polinomio de m-ésimo grado como sigue

$$y = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m + e$$

El análisis anterior se puede extender fácilmente a este caso más general. Así, se reconoce que la determinación de los coeficientes de un polinomio de m-ésimo grado es equivalente a resolver un sistema de $m + 1$ ecuaciones lineales simultáneas. En este caso, el error estándar se formula como sigue:

$$s_{y/x} = \sqrt{\frac{S_r}{n-(m+1)}} \tag{17.20}$$

Esta cantidad se dividide entre $n - (m + 1)$, ya que $(m + 1)$ coeficientes obtenidos de los datos, a_0, $a_1, \ldots, a_m$, se utilizaron para calcular S_r; hemos perdido $m + 1$ grados de libertad. Además del error estándar, también se calcula un coeficiente de determinación para la regresión polinomial con la ecuación (17.10).

EJEMPLO 17.5 Regresión polinomial

Planteamiento del problema Ajustar a un polinomio de segundo grado los datos dados en las dos primeras columnas de la tabla 17.4.

Solución A partir de los datos dados,

$$m = 2 \qquad \sum x_i = 15 \qquad \sum x_i^4 = 979$$
$$n = 6 \qquad \sum y_i = 152.6 \qquad \sum x_i y_i = 585.6$$
$$\bar{x} = 2.5 \qquad \sum x_i^2 = 55 \qquad \sum x_i^2 y_i = 2\,488.8$$
$$\bar{y} = 25.433 \qquad \sum x_i^3 = 225$$

Tabla 17.4 Cálculos para un análisis de error del ajuste cuadrático por mínimos cuadrados.

x_i	y_i	$(y_i - \bar{y})^2$	$(y_i - a_0 - a_1x_i - a_2x_i^2)^2$
0	2.1	544.44	0.14332
1	7.7	314.47	1.00286
2	13.6	140.03	1.08158
3	27.2	3.12	0.80491
4	40.9	239.22	0.61951
5	61.1	1 272.11	0.09439
Σ	152.6	2 513.39	3.74657

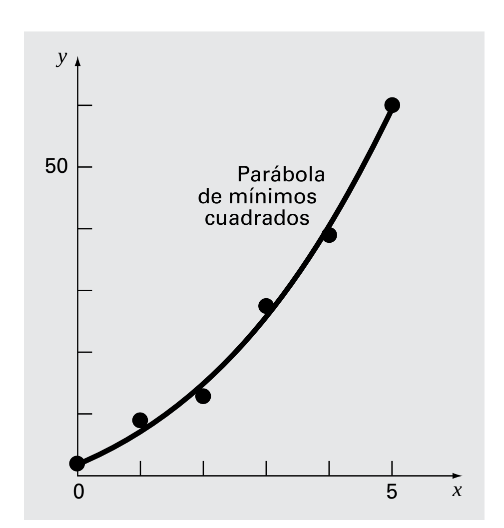


Figura 17.11
Ajuste de un polinomio de segundo grado.

Entonces, las ecuaciones lineales simultáneas son

$$\begin{bmatrix} 6 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} 152.6 \\ 585.6 \\ 2\,488.8 \end{Bmatrix}$$

Resolviendo estas ecuaciones con una técnica como la eliminación de Gauss se tiene $a_0 = 2.47857$, $a_1 = 2.35929$ y $a_2 = 1.86071$. Por lo tanto, la ecuación cuadrática por mínimos cuadrados en este caso es

$$y = 2.47857 + 2.35929x + 1.86071x^2$$

El error estándar de la estimación con base en la regresión polinomial es [ecuación (17.20)]

$$s_{y/x} = \sqrt{\frac{3.74657}{6-3}} = 1.12$$

El coeficiente de determinación es

$$r^2 = \frac{2\,513.39 - 3.74657}{2\,513.39} = 0.99851$$

y el coeficiente de correlación es $r = 0.99925$.

Estos resultados indican que con el modelo se explicó el 99.851% de la incertidumbre original. Este resultado apoya la conclusión de que la ecuación cuadrática representa un excelente ajuste, como también es evidente en la figura 17.11.

17.2.1 Algoritmo para la regresión polinomial

Un algoritmo para la regresión polinomial se expone en la figura 17.12. Observe que la principal tarea es la generación de los coeficientes de las ecuaciones normales [ecuación (17.19)]. (El pseudocódigo para esto se presenta en la figura 17.13.) Las técnicas de la parte tres sirven para resolver estas ecuaciones simultáneas que determinan los coeficientes.

Paso 1: Introduzca el grado del polinomio sujeto a ajuste, m.
Paso 2: Introduzca el número de datos, n.
Paso 3: Si $n < m + 1$, imprima un mensaje de error que indique que la regresión no es posible y termine el proceso. Si $n \geq m + 1$, continúe.
Paso 4: Calcule los elementos de la ecuación normal en la forma de una matriz aumentada.
Paso 5: Usando la matriz aumentada determine los coeficientes $a_0, a_1, a_2, \ldots, a_m$, por medio de un método de eliminación.
Paso 6: Imprima los coeficientes.

Figura 17.12
Algoritmo para implementar la regresión polinomial y lineal múltiple.

Un problema potencial en la implementación de la regresión polinomial en la computadora es que las ecuaciones normales algunas veces están mal condicionadas. Esto se presenta especialmente cuando se plantean polinomios de grado superior. En tales casos, los coeficientes calculados pueden ser altamente susceptibles al error de redondeo y, en consecuencia, los resultados serían inexactos. Entre otras cuestiones, este problema se relaciona con la estructura de las ecuaciones normales y con el hecho de que con polinomios de grado superior las ecuaciones normales pueden tener coeficientes muy grandes y muy pequeños. Lo anterior se debe a que los coeficientes son sumas de datos elevados a potencias.

Aunque las estrategias para disminuir el error de redondeo analizadas en la parte tres, como el pivoteo, pueden ayudar a resolver parcialmente dicho problema, una alternativa más simple consiste en usar una computadora con alta precisión. Por fortuna, la mayoría de los problemas prácticos están limitados a polinomios de grado inferior, en los cuales el error de redondeo generalmente es insignificante. En situaciones donde se requieren versiones de grado superior, se dispone de otras alternativas para ciertos tipos de datos. Sin embargo, esas técnicas (como polinomios ortogonales) están más allá del alcance de este libro. El lector deberá consultar textos sobre regresión, como el de Draper y Smith (1981), para mayor información respecto al problema y sus posibles alternativas.

```
DOFOR i = 1, order + 1
  DOFOR j = 1, i
    k = i + j - 2
    sum = 0
    DOFOR ℓ = 1, n
      sum = sum + x_ℓ^k
    END DO
    a_i,j = sum
    a_j,i = sum
  END DO
  sum = 0
  DOFOR ℓ = 1, n
    sum = sum + y_ℓ · x_ℓ^(i-1)
  END DO
  a_i,order+2 = sum
END DO
```

Figura 17.13
Pseudocódigo para encontrar los elementos de las ecuaciones normales en la regresión polinomial.

17.3 REGRESIÓN LINEAL MÚLTIPLE

Una extensión útil de la regresión lineal es el caso en el que y es una función lineal de dos o más variables independientes. Por ejemplo, y podría ser una función lineal de x_1 y x_2, como en

$$y = a_0 + a_1x_1 + a_2x_2 + e$$

En particular tal ecuación es útil cuando se ajustan datos experimentales donde la variable sujeta a estudio es una función de otras dos variables. En este caso bidimensional, la “línea” de regresión se convierte en un “plano” (figura 17.14).

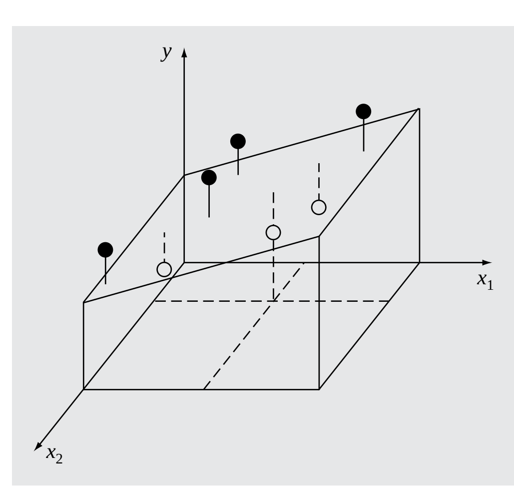


Figura 17.14
Descripción gráfica de una regresión lineal múltiple donde y es una función lineal de x_1 y x_2.

Como en los casos anteriores, los “mejores” valores para los coeficientes se determinan al realizar la suma de los cuadrados de los residuos,

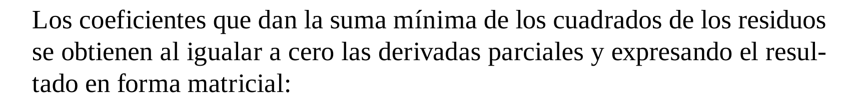
$$S_r = \sum_{i=1}^{n} (y_i - a_0 - a_1x_{1i} - a_2x_{2i})^2 \tag{17.21}$$

y derivando con respecto a cada uno de los coeficientes desconocidos,

$$\frac{\partial S_r}{\partial a_0} = -2\sum (y_i - a_0 - a_1x_{1i} - a_2x_{2i})$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum x_{1i}(y_i - a_0 - a_1x_{1i} - a_2x_{2i})$$

$$\frac{\partial S_r}{\partial a_2} = -2\sum x_{2i}(y_i - a_0 - a_1x_{1i} - a_2x_{2i})$$



Los coeficientes que dan la suma mínima de los cuadrados de los residuos se obtienen al igualar a cero las derivadas parciales y expresando el resultado en forma matricial:

$$\begin{bmatrix} n & \sum x_{1i} & \sum x_{2i} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i}x_{2i} \\ \sum x_{2i} & \sum x_{1i}x_{2i} & \sum x_{2i}^2 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \sum y_i \\ \sum x_{1i}y_i \\ \sum x_{2i}y_i \end{Bmatrix} \tag{17.22}$$

EJEMPLO 17.6 Regresión lineal múltiple

Planteamiento del problema Los siguientes datos se calcularon con la ecuación $y = 5 + 4x_1 - 3x_2$:

x_1	x_2	y
0	0	5
2	1	10
2.5	2	9
1	3	0
4	6	3
7	2	27

Utilice la regresión lineal múltiple para ajustar estos datos.

Solución Las sumatorias requeridas para la ecuación (17.22) se calculan en la tabla 17.5. El resultado es

$$\begin{bmatrix} 6 & 16.5 & 14 \\ 16.5 & 76.25 & 48 \\ 14 & 48 & 54 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} 54 \\ 243.5 \\ 100 \end{Bmatrix}$$

que se resuelve mediante un método como el de eliminación de Gauss, obteniéndose

$$a_0 = 5 \qquad a_1 = 4 \qquad a_2 = -3$$

que es consistente con la ecuación original, de la cual se obtienen los datos.

Tabla 17.5 Cálculos requeridos para desarrollar las ecuaciones normales para el ejemplo 17.6.

	y	x_1	x_2	x_1^2	x_2^2	x_1x_2	x_1y	x_2y
	5	0	0	0	0	0	0	0
	10	2	1	4	1	2	20	10
	9	2.5	2	6.25	4	5	22.5	18
	0	1	3	1	9	3	0	0
	3	4	6	16	36	24	12	18
	27	7	2	49	4	14	189	54
Σ	54	16.5	14	76.25	54	48	243.5	100

El caso bidimensional anterior fácilmente se extiende a m dimensiones así

$$y = a_0 + a_1x_1 + a_2x_2 + \cdots + a_mx_m + e$$

donde el error estándar se formula como

$$s_{y/x} = \sqrt{\frac{S_r}{n-(m+1)}}$$

```
DOFOR i = 1, order + 1
  DOFOR j = 1, i
    sum = 0
    DOFOR ℓ = 1, n
      sum = sum + x_{i−1,ℓ} · x_{j−1,ℓ}
    END DO
    a_{i,j} = sum
    a_{j,i} = sum
  END DO
  sum = 0
  DOFOR ℓ = 1, n
    sum = sum + y_ℓ · x_{i−1,ℓ}
  END DO
  a_{i,order+2} = sum
END DO
```

Figura 17.15
Pseudocódigo para establecer los elementos de las ecuaciones normales en la regresión múltiple. Observe que además de guardar las variables independientes en $x_{1,i}$, $x_{2,i}$, etc., se deben guardar 1 en $x_{0,i}$ para que funcione este algoritmo.

y el coeficiente de determinación se calcula como en la ecuación (17.10). En la figura 17.15 se da un algoritmo para establecer las ecuaciones normales.

Aunque puede haber ciertos casos donde una variable esté linealmente relacionada con dos o más variables, la regresión lineal múltiple tiene además utilidad en la obtención de ecuaciones de potencias de la forma general

$$y = a_0x_1^{a_1}x_2^{a_2} \cdots x_m^{a_m}$$

Tales ecuaciones son extremadamente útiles cuando se ajustan datos experimentales. Para usar regresión lineal múltiple, la ecuación se transforma al aplicar logaritmos:

$$\log y = \log a_0 + a_1 \log x_1 + a_2 \log x_2 + \cdots + a_m \log x_m$$

Esta transformación es similar a la que se usó en la sección 17.1.5 y en el ejemplo 17.4 para ajustar una ecuación de potencias cuando y era una función de una sola variable x. La sección 20.4 muestra un ejemplo de una de estas aplicaciones para dos variables independientes.

17.4 MÍNIMOS CUADRADOS LINEALES EN GENERAL

Hasta aquí nos hemos concentrado en la mecánica para obtener ajustes por mínimos cuadrados de algunas funciones sencillas para datos dados. Antes de ocuparnos de la regresión no lineal, hay varios puntos que nos gustaría analizar para enriquecer nuestra comprensión del material precedente.

17.4.1 Formulación general de una matriz para mínimos cuadrados lineales

En las páginas anteriores presentamos tres tipos de regresión: lineal simple, polinomial y lineal múltiple. De hecho, las tres pertenecen al siguiente modelo lineal general de mínimos cuadrados:

$$y = a_0 z_0 + a_1 z_1 + a_2 z_2 + \cdots + a_m z_m + e \tag{17.23}$$

donde $z_0, z_1, \ldots, z_m$ son $m + 1$ funciones diferentes. Se observa con facilidad cómo la regresión lineal simple y múltiple se encuentran dentro de este modelo; es decir, $z_0 = 1$, $z_1 = x_1$, $z_2 = x_2$, …, $z_m = x_m$. Además, la regresión polinomial se incluye también si las z son monomios simples como $z_0 = x^0 = 1$, $z_1 = x$, $z_2 = x^2, \ldots, z_m = x^m$.

Observe que la terminología "lineal" se refiere sólo a la dependencia del modelo sobre sus parámetros (es decir, las a). Como en el caso de la regresión polinomial, las mismas funciones llegan a ser altamente no lineales. Por ejemplo, las z pueden ser senoidales, como en

$$y = a_0 + a_1 \cos(\omega t) + a_2 \operatorname{sen}(\omega t)$$

Esta forma es la base del análisis de Fourier que se describe en el capítulo 19.

Por otro lado, un modelo de apariencia simple como

$$f(x) = a_0 (1 - e^{-a_1 x})$$

es no lineal porque no es posible llevarlo a la forma de la ecuación (17.23). Regresaremos a tales modelos al final de este capítulo.

Mientras tanto, la ecuación (17.23) se expresa en notación matricial como

$$\{Y\} = [Z]\{A\} + \{E\} \tag{17.24}$$

donde $[Z]$ es una matriz de los valores calculados de las funciones z en los valores medidos de las variables independientes,

$$[Z] = \begin{bmatrix} z_{01} & z_{11} & \cdots & z_{m1} \\ z_{02} & z_{12} & \cdots & z_{m2} \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ z_{0n} & z_{1n} & \cdots & z_{mn} \end{bmatrix}$$

donde m es el número de variables en el modelo y n es el número de datos. Como $n \geq m + 1$, usted reconocerá que, la mayoría de las veces, $[Z]$ no es una matriz cuadrada.

El vector columna $\{Y\}$ contiene los valores observados de la variable dependiente

$$\{Y\}^T = \lfloor y_1 \quad y_2 \quad \cdots \quad y_n \rfloor$$

El vector columna $\{A\}$ contiene los coeficientes desconocidos

$$\{A\}^T = \lfloor a_0 \quad a_1 \quad \cdots \quad a_m \rfloor$$

y el vector columna $\{E\}$ contiene los residuos

$$\{E\}^T = \lfloor e_1 \quad e_2 \quad \cdots \quad e_n \rfloor$$

Como se dio a lo largo de este capítulo, la suma de los cuadrados de los residuos en este modelo se definen como

$$S_r = \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{m} a_j z_{ji} \right)^2$$

Esta cantidad se minimiza tomando las derivadas parciales con respecto a cada uno de los coeficientes e igualando a cero la ecuación resultante. El resultado de este proceso son las ecuaciones normales, que se expresan en forma matricial como

$$[[Z]^T[Z]]\{A\} = \{[Z]^T\{Y\}\} \tag{17.25}$$

Es posible mostrar que la ecuación (17.25) es, de hecho, equivalente a las ecuaciones normales desarrolladas antes para la regresión lineal simple, la polinomial y la múltiple.

Nuestra principal motivación para lo anterior fue ilustrar la unidad entre los tres procedimientos y mostrar cómo se pueden expresar de manera simple en la misma notación matricial. La notación matricial también tendrá relevancia cuando volvamos a la regresión no lineal en la última sección del presente capítulo.

De la ecuación (PT3.6), recuerde que se puede usar la inversa de la matriz para resolver la ecuación (17.25) como en

$$\{A\} = [[Z]^T[Z]]^{-1} = \{[Z]^T\{Y\}\} \tag{17.26}$$

Como se aprendió en la parte tres, éste es un enfoque ineficiente para resolver un sistema de ecuaciones simultáneas. Sin embargo, desde la perspectiva estadística, hay varias razones por las que podría interesar obtener la inversa y examinar sus coeficientes. Estas razones se explicarán en seguida.

17.4.2 Aspectos estadísticos de la teoría de mínimos cuadrados

En la sección PT5.2.1 revisamos diversos estadísticos descriptivos que se utilizan para describir una muestra. Éstos son: la media aritmética, la desviación estándar y la varianza.

Además de dar una solución para los coeficientes de regresión, la formulación matricial de la ecuación (17.26) proporciona estimaciones de sus estadísticos. Es posible demostrar (Draper y Smith, 1981) que los términos en la diagonal y fuera de la diagonal de la matriz $[[Z]^T [Z]]^{-1}$ dan, respectivamente, las varianzas y las covarianzas[1] de las a. Si los elementos de la diagonal de $[[Z]^T [Z]]^{-1}$ se designa por $z_{i,i}^{-1}$, entonces

$$\text{var}(a_{i-1}) = z_{i,i}^{-1} s_{y/x}^2 \tag{17.27}$$

y

$$\text{cov}(a_{i-1}, a_{j-1}) = z_{i,j}^{-1} s_{y/x}^2 \tag{17.28}$$

Dichos estadísticos poseen varias aplicaciones importantes. Para nuestros actuales propósitos, ilustraremos cómo se utilizan para desarrollar intervalos de confianza para la intersección con el eje y y la pendiente.

Con un procedimiento similar al examinado en la sección PT5.2.3, se demuestra que los límites inferior y superior para la intersección con el eje y se pueden encontrar (véase Milton y Arnold, 1995, para más detalles) de la siguiente manera:

$$L = a_0 - t_{\alpha/2,n-2}s(a_0) \qquad U = a_0 + t_{\alpha/2,n-2}s(a_0) \tag{17.29}$$

---

[1] La covarianza es un estadístico que mide la dependencia de una variable respecto de otra. Así, cov(x, y) indica la dependencia de x y y. Por ejemplo, cov(x, y) = 0 indicaría que x y y son totalmente independientes.

donde $s(a_j)$ = el error estándar del coeficiente $a_j = \sqrt{\text{var}(a_j)}$. De manera similar, los límites inferior y superior para la pendiente se calculan:

$$L = a_1 - t_{\alpha/2,n-2}s(a_1) \qquad U = a_1 + t_{\alpha/2,n-2}s(a_1) \tag{17.30}$$

El ejemplo 17.7 ilustra cómo se emplean esos intervalos para realizar inferencias cuantitativas respecto a la regresión lineal.

EJEMPLO 17.7 Intervalos de confianza para la regresión lineal

Planteamiento del problema En el ejemplo 17.3 utilizamos la regresión para desarrollar la siguiente relación entre mediciones y predicciones del modelo:

$$y = -0.859 + 1.032x$$

donde y = las predicciones del modelo y x = las mediciones. Concluimos que había una buena concordancia entre las dos, puesto que la intersección con el eje y era aproximadamente igual a 0, y la pendiente aproximadamente igual a 1. Vuelva a calcular la regresión, pero ahora use el método matricial para estimar los errores estándar de los parámetros. Después emplee tales errores para desarrollar los intervalos de confianza y úselos para realizar un planteamiento probabilístico respecto a la bondad del ajuste.

Solución Los datos se escriben en forma matricial para una regresión lineal simple de la siguiente manera:

$$[Z] = \begin{bmatrix} 1 & 10 \\ 1 & 16.3 \\ 1 & 23 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & 50 \end{bmatrix} \qquad \{Y\} = \begin{Bmatrix} 8.953 \\ 16.405 \\ 22.607 \\ \cdot \\ \cdot \\ \cdot \\ 49.988 \end{Bmatrix}$$

Después se usan la transposición y la multiplicación matriciales para generar las ecuaciones normales:

$$\begin{matrix} [[Z]^T[Z]] & \{A\} & = \{[Z]^T\{Y\}\} \\ \begin{bmatrix} 15 & 548.3 \\ 548.3 & 22\,191.21 \end{bmatrix} & \begin{Bmatrix} a_0 \\ a_1 \end{Bmatrix} & = \begin{Bmatrix} 552.741 \\ 22\,421.43 \end{Bmatrix} \end{matrix}$$

Se emplea la inversión matricial para obtener la pendiente y la intersección con el eje y

$$\begin{aligned} \{A\} &= \qquad [[Z]^T[Z]]^{-1} \qquad \{[Z]^T\{Y\}\} \\ &= \begin{bmatrix} 0.688414 & -0.01701 \\ -0.01701 & 0.000465 \end{bmatrix} \begin{Bmatrix} 552.741 \\ 22\,421.43 \end{Bmatrix} = \begin{Bmatrix} -0.85872 \\ 1.031592 \end{Bmatrix} \end{aligned}$$

De esta manera, la intersección con el eje y y la pendiente quedan como $a_0 = -0.85872$ y $a_1 = 1.031592$, respectivamente. Estos valores, a su vez, sirven para calcular el error estándar del estimado, $s_{y/x} = 0.863403$. Este valor puede utilizarse, junto con los elementos diagonales de la matriz inversa, para calcular los errores estándar de los coeficientes,

$$s(a_0) = \sqrt{z_{11}^{-1}s_{y/x}^2} = \sqrt{0.688414(0.863403)^2} = 0.716372$$

$$s(a_1) = \sqrt{z_{22}^{-1}s_{y/x}^2} = \sqrt{0.000465(0.863403)^2} = 0.018625$$

El estadístico $t_{\alpha/2,n-1}$ necesario para un intervalo de confianza del 95% con $n - 2 = 15 - 2 = 13$ grados de libertad se obtiene con una tabla estadística o mediante software. Usemos una función de Excel, TINV, para obtener el valor adecuado de la siguiente manera:

= TINV(0.05, 13)

que da un valor de 2.160368. Las ecuaciones (17.29) y (17.30) entonces se usan para calcular los intervalos de confianza:

$$a_0 = -0.85872 \pm 2.160368(0.716372)$$
$$= -0.85872 \pm 1.547627 = [-2.40634,\ 0.688912]$$
$$a_1 = 1.031592 \pm 2.160368(0.018625)$$
$$= 1.031592 \pm 0.040237 = [0.991355,\ 1.071828]$$

Observe que los valores deseados (0 para la intersección y la pendiente y 1 para la intersección) caen dentro de los intervalos. Considerando este análisis podremos formular las siguientes declaraciones sobre la pendiente: tenemos fundamentos sólidos para creer que la pendiente de la línea de regresión real está dentro del intervalo de 0.991355 a 1.071828. Debido a que 1 está dentro de este intervalo, también tenemos fundamentos sólidos para creer que el resultado apoya la concordancia entre las mediciones y el modelo. Como cero está dentro del intervalo de la intersección, se puede hacer una declaración similar respecto a la intersección.

Como se mencionó antes en la sección 17.2.1, las ecuaciones normales son notoriamente inadecuadas. Por lo tanto, si se resuelven con técnicas convencionales como la descomposición de *LU*, los coeficientes calculados pueden ser muy propensos a errores de redondeo. En consecuencia, para darle la vuelta al problema, hay disponibles algoritmos más tecnificados de ortogonalización como la *factorización QR*. Como estas técnicas están más allá del alcance de este libro, el lector debe consultar textos sobre regresión, como Draper y Smith (1981) para información adicional acerca del problema y posibles alternativas. Moler (2004) también ofrece una buena explicación del tema, poniendo énfasis en los métodos numéricos.

Lo anterior es una introducción limitada al rico tema de la inferencia estadística y su relación con la regresión. Hay muchas sutilezas que rebasan el alcance de este libro, cuya motivación principal ha sido ilustrar la potencia del enfoque matricial a los mínimos cuadrados lineales generales. Además, se debe observar que paquetes de software como Excel, MATLAB y Mathcad pueden generar ajuste de datos de regresión por mínimos cuadrados junto con información de importancia para la inferencia estadística. Se explorarán algunas de estas capacidades cuando se describan estos paquetes al final del capítulo 19.

17.5 REGRESIÓN NO LINEAL

Hay muchos casos en la ingeniería donde los modelos no lineales deben ajustarse a datos. En el presente contexto, tales modelos se definen como aquellos que tienen dependencia no lineal de sus parámetros. Por ejemplo,

$$f(x) = a_0(1 - e^{-a_1 x}) + e \tag{17.31}$$

Esta ecuación no puede ser manipulada para ser llevada a la forma general de la ecuación (17.23).

Como en el caso de los mínimos cuadrados lineales, la regresión no lineal se basa en la determinación de los valores de los parámetros que minimizan la suma de los cuadrados de los residuos. Sin embargo, en el caso no lineal, la solución debe realizarse en una forma iterativa.

El *método de Gauss-Newton* es un algoritmo para minimizar la suma de los cuadrados de los residuos entre los datos y las ecuaciones no lineales. El concepto clave detrás de esta técnica es que se utiliza una expansión en serie de Taylor para expresar la ecuación no lineal original en una forma lineal aproximada. Entonces, es posible aplicar la teoría de mínimos cuadrados para obtener nuevas estimaciones de los parámetros que se mueven en la dirección que minimiza el residuo.

Para ilustrar cómo se logra esto, primero se expresa de manera general la relación entre la ecuación no lineal y los datos, de la manera siguiente:

$$y_i = f(x_i;\, a_0, a_1, \ldots, a_m) + e_i$$

donde y_i = un valor medido de la variable dependiente, $f(x_i; a_0, a_1, \ldots, a_m)$ = la ecuación que es una función de la variable independiente x_i y una función no lineal de los parámetros $a_0, a_1, \ldots, a_m$, y e_i = un error aleatorio. Por conveniencia, este modelo se expresa en forma abreviada al omitir los parámetros,

$$y_i = f(x_i) + e_i \tag{17.32}$$

El modelo no lineal puede expandirse en una serie de Taylor alrededor de los valores de los parámetros y cortarse después de las primeras derivadas. Por ejemplo, para un caso con dos parámetros,

$$f(x_i)_{j+1} = f(x_i)_j + \frac{\partial f(x_i)_j}{\partial a_0}\Delta a_0 + \frac{\partial f(x_i)_j}{\partial a_1}\Delta a_1 \tag{17.33}$$

donde j = el valor inicial, $j + 1$ = la predicción, $\Delta a_0 = a_{0,j+1} - a_{0,j}$, y $\Delta a_1 = a_{1,j+1} - a_{1,j}$. De esta forma, hemos linealizado el modelo original con respecto a los parámetros. La ecuación (17.33) se sustituye en la ecuación (17.32) para dar

$$y_i - f(x_i)_j = \frac{\partial f(x_i)_j}{\partial a_0}\Delta a_0 + \frac{\partial f(x_i)_j}{\partial a_1}\Delta a_1 + e_i$$

o en forma matricial [compárela con la ecuación (17.24)],

$$\{D\} = [Z_j]\{\Delta A\} + \{E\} \tag{17.34}$$

donde $[Z_j]$ es la matriz de las derivadas parciales de la función evaluadas en el valor inicial j,

$$[Z_j] = \begin{bmatrix} \partial f_1/\partial a_0 & \partial f_1/\partial a_1 \\ \partial f_2/\partial a_0 & \partial f_2/\partial a_1 \\ . & . \\ . & . \\ . & . \\ \partial f_n/\partial a_0 & \partial f_n/\partial a_1 \end{bmatrix}$$

donde n = el número de datos y $\partial f_i/\partial a_k$ = la derivada parcial de la función con respecto al k-ésimo parámetro evaluado en el i-ésimo dato. El vector $\{D\}$ contiene las diferencias entre las mediciones y los valores de la función,

$$\{D\} = \begin{Bmatrix} y_1 - f(x_1) \\ y_2 - f(x_2) \\ . \\ . \\ . \\ y_n - f(x_n) \end{Bmatrix}$$

y el vector $\{\Delta A\}$ contiene los cambios en los valores de los parámetros,

$$\{\Delta A\} = \begin{Bmatrix} \Delta a_0 \\ \Delta a_1 \\ . \\ . \\ . \\ \Delta a_m \end{Bmatrix}$$

Si se aplica la teoría de los mínimos cuadrados lineales a la ecuación (17.34) se obtienen las siguientes ecuaciones normales [recuerde la ecuación (17.25)]:

$$[[Z_j]^T[Z_j]]\{\Delta A\} = \{[Z_j]^T\{D\}\} \tag{17.35}$$

Así, el procedimiento consiste en resolver de la ecuación (17.35) para $\{\Delta A\}$, que se utiliza para calcular valores mejorados de los parámetros, como en

$$a_{0,j+1} = a_{0,j} + \Delta a_0$$

y

$$a_{1,j+1} = a_{1,j} + \Delta a_1$$

Este procedimiento se repite hasta que la solución converge, es decir, hasta que

$$|\varepsilon_a|_k = \left|\frac{a_{k,j+1} - a_{k,j}}{a_{k,j+1}}\right| 100\% \tag{17.36}$$

está por debajo de un criterio de terminación aceptable.

EJEMPLO 17.8 Método de Gauss-Newton

Planteamiento del problema Ajuste la función $f(x; a_0, a_1) = a_0(1 - e^{-a_1 x})$ a los datos:

x	0.25	0.75	1.25	1.75	2.25
y	0.28	0.57	0.68	0.74	0.79

Emplee $a_0 = 1.0$ y $a_1 = 1.0$ como valores iniciales para los parámetros. Observe que para estos valores la suma inicial de los cuadrados de los residuos es 0.0248.

Solución Las derivadas parciales de la función con respecto a los parámetros son

$$\frac{\partial f}{\partial a_0} = 1 - e^{-a_1 x} \tag{E17.8.1}$$

y

$$\frac{\partial f}{\partial a_1} = a_0 x e^{-a_1 x} \tag{E17.8.2}$$

Las ecuaciones (E17.8.1) y (E17.8.2) se utilizan para evaluar la matriz

$$[Z_0] = \begin{bmatrix} 0.2212 & 0.1947 \\ 0.5276 & 0.3543 \\ 0.7135 & 0.3581 \\ 0.8262 & 0.3041 \\ 0.8946 & 0.2371 \end{bmatrix}$$

Esta matriz multiplicada por su transpuesta nos da

$$[Z_0]^T[Z_0] = \begin{bmatrix} 2.3193 & 0.9489 \\ 0.9489 & 0.4404 \end{bmatrix}$$

la cual, a su vez, se invierte con el siguiente resultado:

$$[[Z_0]^T[Z_0]]^{-1} = \begin{bmatrix} 3.6397 & -7.8421 \\ -7.8421 & 19.1678 \end{bmatrix}$$

El vector $\{D\}$ consiste en las diferencias entre las mediciones y las predicciones del modelo,

$$\{D\} = \begin{Bmatrix} 0.28 - 0.2212 \\ 0.57 - 0.5276 \\ 0.68 - 0.7135 \\ 0.74 - 0.8262 \\ 0.79 - 0.8946 \end{Bmatrix} = \begin{Bmatrix} 0.0588 \\ 0.0424 \\ -0.0335 \\ -0.0862 \\ -0.1046 \end{Bmatrix}$$

Éste se multiplica por $[Z_0]^T$ para dar

$$[Z_0]^T\{D\} = \begin{bmatrix} -0.1533 \\ -0.0365 \end{bmatrix}$$

El vector $\{\Delta A\}$, entonces, se calcula al resolver la ecuación (17.35):

$$\Delta A = \begin{Bmatrix} -0.2714 \\ 0.5019 \end{Bmatrix}$$

que se suma a los valores iniciales de los parámetros:

$$\begin{Bmatrix} a_0 \\ a_1 \end{Bmatrix} = \begin{Bmatrix} 1.0 \\ 1.0 \end{Bmatrix} + \begin{Bmatrix} -0.2714 \\ 0.5019 \end{Bmatrix} = \begin{Bmatrix} 0.7286 \\ 1.5019 \end{Bmatrix}$$

Así, los estimados mejorados de los parámetros son $a_0 = 0.7286$ y $a_1 = 1.5019$. Los nuevos parámetros dan una suma de los cuadrados de los residuos igual a 0.0242. La ecuación (17.36) se utiliza para obtener que ε_0 y ε_1 son iguales a 37 y 33%, respectivamente. El cálculo se repetiría hasta que esos valores estén abajo del criterio de terminación establecido. El resultado final es $a_0 = 0.79186$ y $a_1 = 1.6751$. Tales coeficientes dan una suma de los cuadrados de los residuos de 0.000662.

Un problema potencial con el método de Gauss-Newton, como se ha desarrollado hasta ahora, es que las derivadas parciales de la función pueden ser difíciles de evaluar. En consecuencia, muchos programas computacionales usan diferentes ecuaciones para aproximar las derivadas parciales. Un método es

$$\frac{\partial f_i}{\partial a_k} \cong \frac{f(x_i; a_0, \ldots, a_k + \delta a_k, \ldots, a_m) - f(x_i; a_0 \ldots, a_k, \ldots, a_m)}{\delta a_k} \tag{17.37}$$

donde δ = una perturbación fraccional pequeña.

El método de Gauss-Newton tiene también algunas desventajas:

1. Puede converger con lentitud.
2. Puede oscilar ampliamente; es decir, cambia de dirección continuamente.
3. Puede no converger.

Se han desarrollado modificaciones del método (Booth y Peterson, 1958; Hartley, 1961) para disminuir las desventajas.

Además, aunque hay varios procedimientos expresamente diseñados para regresión, un método más general es usar rutinas de optimización no lineal como las descritas en la parte cuatro. Para hacer esto, se dan valores iniciales a los parámetros y se calcula la suma de los cuadrados de los residuos. Por ejemplo, para la ecuación (17.31) esto se podría calcular como

$$S_r = \sum_{i=1}^{n} [y_i - a_0(1 - e^{-a_1 x_i})]^2 \tag{17.38}$$

Los parámetros, entonces, se ajustarían de manera sistemática para minimizar S_r mediante técnicas de búsqueda como las descritas previamente en el capítulo 14. Ilustraremos el modo para hacer esto cuando describamos las aplicaciones de software, al final del capítulo 19.

PROBLEMAS

17.1 Dados los siguientes datos

8.8	9.5	9.8	9.4	10.0
9.4	10.1	9.2	11.3	9.4
10.0	10.4	7.9	10.4	9.8
9.8	9.5	8.9	8.8	10.6
10.1	9.5	9.6	10.2	8.9

determine *a*) la media, *b*) la desviación estándar, *c*) la varianza, *d*) el coeficiente de variación y *e*) el intervalo de confianza de 95% para la media. *f*) Construya un histograma usando un rango de 7.5 a 11.5 con intervalos de 0.5.

17.2 Dados los datos

29.65	28.55	28.65	30.15	29.35	29.75	29.25
30.65	28.15	29.85	29.05	30.25	30.85	28.75
29.65	30.45	29.15	30.45	33.65	29.35	29.75
31.25	29.45	30.15	29.65	30.55	29.65	29.25

Determine *a*) la media, *b*) la desviación estándar, *c*) la varianza, *d*) el coeficiente de variación y *e*) el intervalo de confianza del 90% para la media. *f*) Construya un histograma. Use un rango de 28 a 34 con incrementos de 0.4. *g*) Si se supone que la distribución es normal y que la estimación de la desviación estándar es válida, calcule el rango (es decir, los valores inferior y superior) que agrupa al 68% de los datos. Determine si ésta es una estimación válida para los datos del problema.

17.3 Utilice la regresión por mínimos cuadrados para ajustar una línea recta a

x	0	2	4	6	9	11	12	15	17	19
y	5	6	7	6	9	8	7	10	12	12

Además de la pendiente y la intersección, calcule el error estándar de la estimación y el coeficiente de correlación. Haga una gráfica de los datos y la línea de regresión. Después repita el problema, pero ahora efectúe la regresión de *x versus y*, es decir, intercambie las variables. Interprete sus resultados.

17.4 Use la regresión por mínimos cuadrados para ajustar una línea recta a

x	6	7	11	15	17	21	23	29	29	37	39
y	29	21	29	14	21	15	7	7	13	0	3

Además de la pendiente y la intersección, calcule el error estándar de la estimación y el coeficiente de correlación. Haga una gráfica de los datos y la línea de regresión. ¿Si otra persona hiciera una medición adicional de $x = 10$, $y = 10$, usted pensaría, con base en una evaluación visual y el error estándar, que la medición era válida o inválida? Justifique su conclusión.

17.5 Con el mismo enfoque que se empleó para obtener las ecuaciones (17.15) y (17.16), obtenga el ajuste por mínimos cuadrados del modelo siguiente:

$$y = a_1x + e$$

Es decir, determine la pendiente que resulta en el ajuste por mínimos cuadrados para una línea recta con intersección en el origen. Ajuste los datos siguientes con dicho modelo e ilustre el resultado con una gráfica:

x	2	4	6	7	10	11	14	17	20
y	1	2	5	2	8	7	6	9	12

17.6 Emplee la regresión por mínimos cuadrados para ajustar una línea recta a

x	1	2	3	4	5	6	7	8	9
y	1	1.5	2	3	4	5	8	10	13

a) Además de la pendiente y la intersección, calcule el error estándar de la estimación y el coeficiente de correlación. Grafique los datos y la línea recta. Evalúe el ajuste.

b) Vuelva a hacer el cálculo del inciso *a*), pero use regresión polinomial para ajustar una parábola a los datos. Compare los resultados con los del inciso *a*).

17.7 Ajuste los datos siguientes con *a*) un modelo de tasa de crecimiento de saturación, *b*) una ecuación de potencias y *c*) una parábola. En cada caso, haga una gráfica de los datos y la ecuación.

x	0.75	2	3	4	6	8	8.5
y	1.2	1.95	2	2.4	2.4	2.7	2.6

17.8 Ajuste los datos siguientes con el modelo de potencias ($y = ax^b$). Use la ecuación de potencias resultante para hacer el pronóstico de y en $x = 9$.

x	2.5	3.5	5	6	7.5	10	12.5	15	17.5	20
y	13	11	8.5	8.2	7	6.2	5.2	4.8	4.6	4.3

17.9 Ajuste a un modelo exponencial a

x	0.4	0.8	1.2	1.6	2	2.3
y	800	975	1 500	1 950	2 900	3 600

Grafique los datos y la ecuación tanto en papel milimétrico como en semilogarítmico.

17.10 En vez de usar el modelo exponencial de base *e* (ecuación 17.22), una alternativa común consiste en utilizar un modelo de base 10.

$$y = \alpha_5 10^{\beta_5 x}$$

Cuando se usa para ajustar curvas, esta ecuación lleva a resultados idénticos que los de la versión con base *e*, pero el valor del parámetro del exponente (β_5) difiere del estimado con la ecuación (17.22) (β_1). Use la versión con base 10 para resolver el problema 17.9. Además, desarrolle una formulación para relacionar β_1 con β_5.

17.11 Además de los ejemplos de la figura 17.10, existen otros modelos que se pueden hacer lineales con el empleo de transformaciones. Por ejemplo,

$$y = \alpha_4 x e^{\beta_4 x}$$

Haga lineal este modelo y úselo para estimar α_4 y β_4 con base en los datos siguientes. Elabore una gráfica del ajuste junto con los datos.

x	0.1	0.2	0.4	0.6	0.9	1.3	1.5	1.7	1.8
y	0.75	1.25	1.45	1.25	0.85	0.55	0.35	0.28	0.18

17.12 Un investigador informa sobre los datos tabulados a continuación de un experimento para determinar la tasa de crecimiento de bacterias *k* (por día), como función de la concentración de oxígeno *c* (mg/L). Se sabe que dichos datos pueden modelarse por medio de la ecuación siguiente:

$$k = \frac{k_{\text{máx}} c^2}{c_s + c^2}$$

donde c_s y $k_{\text{máx}}$ son parámetros. Use una transformación para hacer lineal esta ecuación. Después utilice regresión lineal para estimar c_s y $k_{\text{máx}}$, y pronostique la tasa de crecimiento para $c = 2$ mg/L.

c	0.5	0.8	1.5	2.5	4
k	1.1	2.4	5.3	7.6	8.9

17.13 Un investigador ha reportado los datos que se tabulan en seguida. Se sabe que dichos datos se pueden modelar por la siguiente ecuación:

$$x = e^{(y-b)/a}$$

donde *a* y *b* son parámetros. Use una transformación para linealizar esta ecuación y luego utilice la regresión lineal para determinar *a* y *b*. Con base en su análisis, prediga *y* para $x = 2.6$:

x	1	2	3	5	5
y	0.5	2	2.9	3.5	4

17.14 Se sabe que los datos que se tabulan en seguida se pueden modelar por la siguiente ecuación:

$$y = \left(\frac{a + \sqrt{x}}{b\sqrt{x}}\right)^2$$

Use una transformación para linealizar esta ecuación y luego utilice la regresión lineal para determinar los parámetros *a* y *b*. Con base en su análisis, prediga *y* para $x = 1.6$.

x	0.5	1	2	3	4
y	10.4	5.8	3.3	2.4	2

17.15 Se proporcionan los siguientes datos:

x	1	2	3	4	5
y	2.2	2.8	3.6	4.5	5.5

Usted quiere usar regresión por mínimos cuadrados para ajustar estos datos con el siguiente modelo:

$$y = a + bx + \frac{c}{x}$$

Determine los coeficientes estableciendo y resolviendo la ecuación (17.25).

17.16 Dados los datos

x	5	10	15	20	25	30	35	40	45	50
y	17	24	31	33	37	37	40	40	42	41

use regresión por mínimos cuadrados para ajustar *a*) una línea recta, *b*) una ecuación de potencias, *c*) una ecuación de tasa de crecimiento de saturación, y *d*) una parábola. Grafique los datos junto con todas las curvas. ¿Alguna de las curvas es superior a las demás? Si así fuera, justifíquelo.

17.17 Ajuste una ecuación cúbica a los datos siguientes:

x	3	4	5	7	8	9	11	12
y	1.6	3.6	4.4	3.4	2.2	2.8	3.8	4.6

Además de los coeficientes, determine r^2 y $s_{y/x}$.

17.18 Utilice regresión lineal múltiple para ajustar

x_1	0	1	1	2	2	3	3	4	4
x_2	0	1	2	1	2	1	2	1	2
y	15.1	17.9	12.7	25.6	20.5	35.1	29.7	45.4	40.2

Calcule los coeficientes, el error estándar de la estimación y el coeficiente de correlación.

17.19 Use regresión lineal múltiple para ajustar

x_1	0	0	1	2	0	1	2	2	1
x_2	0	2	2	4	4	6	6	2	1
y	14	21	11	12	23	23	14	6	11

Calcule los coeficientes, el error estándar de la estimación y el coeficiente de correlación.

17.20 Emplee regresión no lineal para ajustar una parábola a los datos siguientes:

x	0.2	0.5	0.8	1.2	1.7	2	2.3
y	500	700	1 000	1 200	2 200	2 650	3 750

17.21 Use regresión no lineal para ajustar una ecuación de tasa de crecimiento de saturación a los datos del problema 17.16.

17.22 Vuelva a calcular los ajustes de regresión de los problemas *a*) 17.3, y *b*) 17.17, con el enfoque matricial. Estime los errores estándar y desarrolle intervalos de confianza del 90% para los coeficientes.

17.23 Desarrolle, depure y pruebe un programa en cualquier lenguaje de alto nivel o de macros que elija, para implantar el análisis de regresión lineal. Entre otras cosas: *a*) incluya comentarios para documentar el código, y *b*) determine el error estándar y el coeficiente de determinación.

17.24 Se hace la prueba a un material para estudiar la falla por fatiga cíclica, en la que se aplica un esfuerzo, en MPa, al material y se mide el número de ciclos que se necesita para hacer que falle. Los resultados se presentan en la tabla siguiente. Al hacerse una gráfica log-log, del esfuerzo *versus* los ciclos, la tendencia de los datos

presenta una relación lineal. Use regresión por mínimos cuadrados para determinar la ecuación de mejor ajuste para dichos datos.

N, ciclos	1	10	100	1 000	10 000	100 000	1 000 000
Esfuerzo, MPa	1 100	1 000	925	800	625	550	420

17.25 Los datos siguientes muestran la relación entre la viscosidad del aceite SAE 70 y su temperatura. Después de obtener el logaritmo de los datos, use regresión lineal para encontrar la ecuación de la recta que se ajuste mejor a los datos y al valor de r^2.

Temperatura, °C	26.67	93.33	148.89	315.56
Viscosidad, μ, N · s/m^2	1.35	0.085	0.012	0.00075

17.26 Los datos siguientes representan el crecimiento bacterial en un cultivo líquido durante cierto número de días.

Día	0	4	8	12	16	20
Cantidad $\times 10^6$	67	84	98	125	149	185

Encuentre la ecuación de mejor ajuste a la tendencia de los datos. Pruebe varias posibilidades: lineal, parabólica y exponencial. Utilice el paquete de software de su elección para obtener la mejor ecuación para pronosticar la cantidad de bacterias después de 40 días.

17.27 Después de una tormenta, se vigila la concentración de la bacteria *E. coli* en un área de natación:

t (h)	4	8	12	16	20	24
c (CFU/100 mL)	1 600	1 320	1 000	890	650	560

El tiempo se mide en horas transcurridas después de finalizar la tormenta, y la unidad CFU es una "unidad de formación de colonia". Use los datos para estimar *a*) la concentración al final de la tormenta ($t = 0$), y *b*) el tiempo en el que la concentración alcanzará 200 CFU/100 mL. Observe que la elección del modelo debe ser consistente con el hecho de que las concentraciones negativas son imposibles y de que la concentración de bacterias siempre disminuye con el tiempo.

17.28 Un objeto se suspende en un túnel de viento y se mide la fuerza para varios niveles de velocidad del viento. A continuación están tabulados los resultados.

v, m/s	10	20	30	40	50	60	70	80
F, N	25	70	380	550	610	1 220	830	1 450

Emplee regresión por mínimos cuadrados para ajustar estos datos con *a*) una línea recta, *b*) una ecuación de potencias basada en transformaciones logarítmicas, y *c*) un modelo de potencias con base en regresión no lineal. Muestre los resultados gráficamente.

17.29 Ajuste un modelo de potencias a los datos del problema 17.28, pero emplee logaritmos naturales para hacer las transformaciones.

17.30 Obtenga el ajuste por mínimos cuadrados del modelo siguiente:

$$y = a_1x + a_2x^2 + e$$

Es decir, determine los coeficientes que generan el ajuste por mínimos cuadrados de un polinomio de segundo grado con intersección en el origen. Pruebe el enfoque con el ajuste de los datos del problema 17.28.

17.31 En el problema 17.11, en el que se usaron transformaciones para hacer lineal y ajustar el modelo siguiente:

$$y = \alpha_4 x e^{\beta_4 x}$$

emplee regresión no lineal para estimar α_4 y β_4 con base en los datos siguientes. Haga una gráfica del ajuste junto con los datos.

x	0.1	0.2	0.4	0.6	0.9	1.3	1.5	1.7	1.8
y	0.75	1.25	1.45	1.25	0.85	0.55	0.35	0.28	0.18

CAPÍTULO 18

Interpolación

Con frecuencia se encontrará con que tiene que estimar valores intermedios entre puntos asociados con datos. El método más común que se usa para este propósito es la interpolación polinomial. Recuerde que la fórmula general para un polinomio de n-ésimo grado es

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \tag{18.1}$$

Dados $n + 1$ puntos asociados con datos, hay uno y sólo un polinomio de grado n que pasa a través de todos los puntos. Por ejemplo, hay sólo una línea recta (es decir, un polinomio de primer grado) que une dos puntos (figura 18.1*a*). De manera similar, únicamente una parábola une un conjunto de tres puntos (figura 18.1*b*). La *interpolación polinomial* consiste en determinar el polinomio único de n-ésimo grado que se ajuste a $n + 1$ puntos asociados con datos. Este polinomio, entonces, proporciona una fórmula para calcular valores intermedios.

Aunque hay uno y sólo un polinomio de n-ésimo grado que se ajusta a $n + 1$ puntos, existe una gran variedad de formas matemáticas en las cuales puede expresarse este polinomio. En este capítulo describiremos dos alternativas que son muy adecuadas para implementarse en computadora: los polinomios de Newton y de Lagrange.

18.1 INTERPOLACIÓN POLINOMIAL DE NEWTON EN DIFERENCIAS DIVIDIDAS

Como se dijo antes, existe una gran variedad de formas alternativas para expresar una interpolación polinomial. El *polinomio de interpolación de Newton en diferencias divididas* es una de las formas más populares y útiles. Antes de presentar la ecuación general, estudiaremos las versiones de primero y segundo grados por su sencilla interpretación visual.

18.1.1 Interpolación lineal

La forma más simple de interpolación consiste en unir dos puntos asociados con datos con una línea recta. Dicha técnica, llamada *interpolación lineal*, se ilustra de manera gráfica en la figura 18.2. Utilizando triángulos semejantes,

$$\frac{f_1(x) - f(x_0)}{x - x_0} = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

reordenándose se tiene

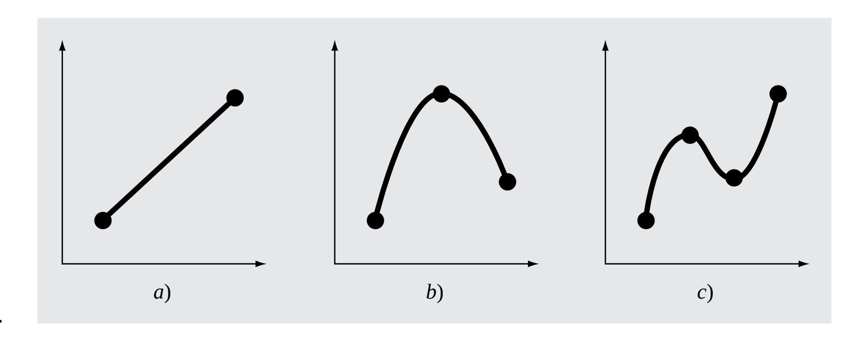


Figura 18.1
Ejemplos de interpolación polinomial: *a*) de primer grado (lineal) que une dos puntos, *b*) de segundo grado (cuadrática o parabólica) que une tres puntos y c) de tercer grado (cúbica) que une cuatro puntos.

$$f_1(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0) \tag{18.2}$$

que es una *fórmula de interpolación lineal*. La notación $f_1(x)$ designa que éste es un polinomio de interpolación de primer grado. Observe que además de representar la pendiente de la línea que une los puntos, el término $[f(x_1) - f(x_0)]/(x_1 - x_0)$ es una aproximación en diferencias divididas finitas a la primera derivada [ecuación (4.17)]. En general, cuanto menor sea el intervalo entre los puntos asociados con datos, mejor será la aproximación. Esto se debe al hecho de que, conforme el intervalo disminuye, una función continua estará mejor aproximada por una línea recta. Esta característica se demuestra en el siguiente ejemplo.

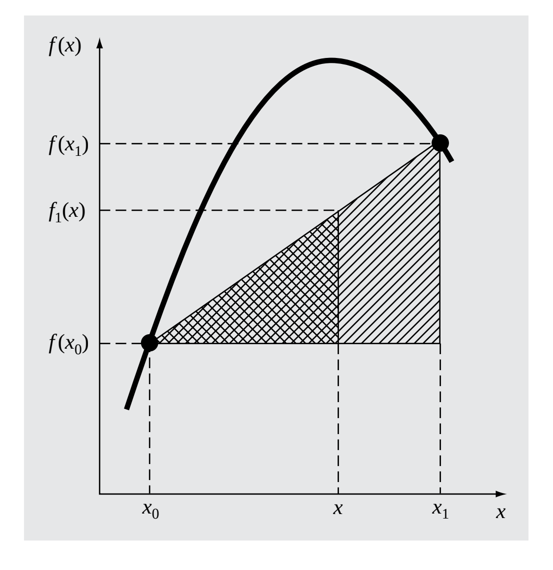


Figura 18.2
Esquema gráfico de la interpolación lineal. Las áreas sombreadas indican los triángulos semejantes usados para obtener la fórmula de la interpolación lineal [ecuación(18.2)].

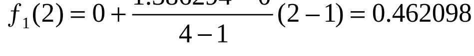
EJEMPLO 18.1 Interpolación lineal

Planteamiento del problema Estime el logaritmo natural de 2 mediante interpolación lineal. Primero, realice el cálculo por interpolación entre ln 1 = 0 y ln 6 = 1.791759. Después, repita el procedimiento, pero use un intervalo menor de ln 1 a ln 4 (1.386294). Observe que el valor verdadero de ln 2 es 0.6931472.

Solución Usamos la ecuación (18.2) y una interpolación lineal para ln(2) desde $x_0 = 1$ hasta $x_1 = 6$ para obtener

$$f_1(2) = 0 + \frac{1.791759 - 0}{6 - 1}(2 - 1) = 0.3583519$$

que representa un error: $\varepsilon_t = 48.3\%$. Con el intervalo menor desde $x_0 = 1$ hasta $x_1 = 4$ se obtiene



$$f_1(2) = 0 + \frac{1.386294 - 0}{4 - 1}(2 - 1) = 0.4620981$$

Así, usando el intervalo más corto el error relativo porcentual se reduce a $\varepsilon_t = 33.3\%$. Ambas interpolaciones se muestran en la figura 18.3, junto con la función verdadera.

18.1.2 Interpolación cuadrática

En el ejemplo 18.1 el error resulta de nuestra aproximación a una curva mediante una línea recta. En consecuencia, una estrategia para mejorar la estimación consiste en introducir alguna curvatura a la línea que une los puntos. Si se tienen tres puntos asociados con datos, éstos pueden ajustarse en un polinomio de segundo grado (también conocido como polinomio cuadrático o *parábola*). Una forma particularmente conveniente para ello es

$$f_2(x) = b_0 + b_1(x - x_0) + b_2(x - x_0)(x - x_1) \tag{18.3}$$

Observe que aunque la ecuación (18.3) parece diferir del polinomio general [ecuación (18.1)], las dos ecuaciones son equivalentes. Lo anterior se demuestra al multiplicar los términos de la ecuación (18.3):

$$f_2(x) = b_0 + b_1x - b_1x_0 + b_2x^2 + b_2x_0x_1 - b_2xx_0 - b_2xx_1$$

o, agrupando términos,

$$f_2(x) = a_0 + a_1x + a_2x^2$$

donde

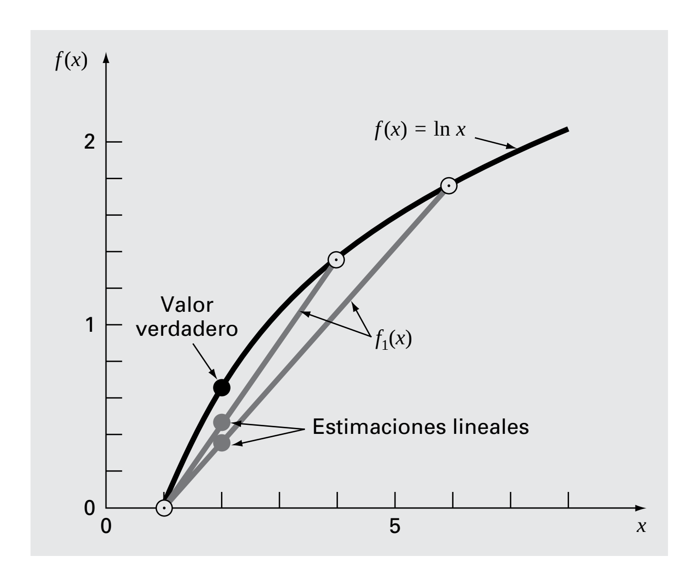


Figura 18.3
Dos interpolaciones lineales para estimar ln 2. Observe cómo el intervalo menor proporciona una mejor estimación.

$$a_0 = b_0 - b_1x_0 + b_2x_0x_1$$
$$a_1 = b_1 - b_2x_0 - b_2x_1$$
$$a_2 = b_2$$

Así, las ecuaciones (18.1) y (18.3) son formas alternativas, equivalentes del único polinomio de segundo grado que une los tres puntos.

Un procedimiento simple puede usarse para determinar los valores de los coeficientes. Para encontrar b_0, en la ecuación (18.3) se evalúa con $x = x_0$ para obtener

$$b_0 = f(x_0) \tag{18.4}$$

La ecuación (18.4) se sustituye en la (18.3), después se evalúa en $x = x_1$ para tener

$$b_1 = \frac{f(x_1) - f(x_0)}{x_1 - x_0} \tag{18.5}$$

Por último, las ecuaciones (18.4) y (18.5) se sustituyen en la (18.3), después se evalúa en $x = x_2$ y (luego de algunas manipulaciones algebraicas) se resuelve para

$$b_2 = \frac{\dfrac{f(x_2) - f(x_1)}{x_2 - x_1} - \dfrac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0} \tag{18.6}$$

Observe que, como en el caso de la interpolación lineal, b_1 todavía representa la pendiente de la línea que une los puntos x_0 y x_1. Así, los primeros dos términos de la ecuación (18.3) son equivalentes a la interpolación lineal de x_0 a x_1, como se especificó antes en la ecuación (18.2). El último término, $b_2(x - x_0)(x - x_1)$, determina la curvatura de segundo grado en la fórmula.

Antes de ilustrar cómo utilizar la ecuación (18.3), debemos examinar la forma del coeficiente b_2. Es muy similar a la aproximación en diferencias divididas finitas de la segunda derivada, que se presentó antes en la ecuación (4.24). Así, la ecuación (18.3) comienza a manifestar una estructura semejante a la expansión de la serie de Taylor. Esta observación será objeto de una mayor exploración cuando relacionemos los polinomios de interpolación de Newton con la serie de Taylor en la sección 18.1.4. Aunque, primero, mostraremos un ejemplo que indique cómo se utiliza la ecuación (18.3) para interpolar entre tres puntos.

EJEMPLO 18.2 Interpolación cuadrática

Planteamiento del problema Ajuste un polinomio de segundo grado a los tres puntos del ejemplo 18.1:

$x_0 = 1$ $\quad$ $f(x_0) = 0$
$x_1 = 4$ $\quad$ $f(x_1) = 1.386294$
$x_2 = 6$ $\quad$ $f(x_2) = 1.791759$

Con el polinomio evalúe ln 2.

Solución Aplicando la ecuación (18.4) se obtiene

$$b_0 = 0$$

La ecuación (18.5) da

$$b_1 = \frac{1.386294 - 0}{4 - 1} = 0.4620981$$

y con la ecuación (18.6) se obtiene

$$b_2 = \frac{\dfrac{1.791759 - 1.386294}{6 - 4} - 0.4620981}{6 - 1} = -0.0518731$$

Sustituyendo estos valores en la ecuación (18.3) se obtiene la fórmula cuadrática

$$f_2(x) = 0 + 0.4620981(x - 1) - 0.0518731(x - 1)(x - 4)$$

que se evalúa en $x = 2$ para

$$f_2(2) = 0.5658444$$

que representa un error relativo de $\varepsilon_t = 18.4\%$. Así, la curvatura determinada por la fórmula cuadrática (figura 18.4) mejora la interpolación comparándola con el resultado obtenido antes al usar las líneas rectas del ejemplo 18.1 y en la figura 18.3.

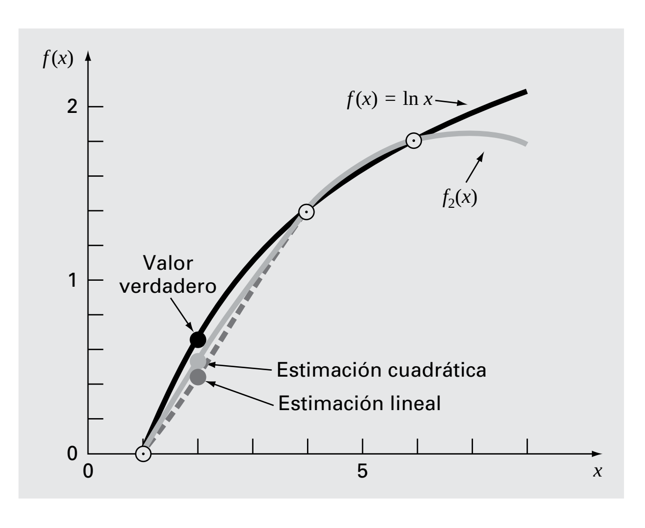


Figura 18.4
El uso de la interpolación cuadrática para estimar ln 2. Para comparación se presenta también la interpolación lineal desde $x = 1$ hasta 4.

18.1.3 Forma general de los polinomios de interpolación de Newton

El análisis anterior puede generalizarse para ajustar un polinomio de n-ésimo grado a $n + 1$ puntos asociados con datos. El polinomio de n-ésimo grado es

$$f_n(x) = b_0 + b_1(x - x_0) + \cdots + b_n(x - x_0)(x - x_1)\cdots(x - x_{n-1}) \tag{18.7}$$

Como se hizo antes con las interpolaciones lineales y cuadráticas, los puntos asociados con datos se utilizan para evaluar los coeficientes $b_0, b_1, \ldots, b_n$. Para un polinomio de n-ésimo grado se requieren $n + 1$ puntos asociados con datos: $[x_0, f(x_0)], [x_1, f(x_1)], \ldots, [x_n, f(x_n)]$. Usamos estos puntos asociados con datos y las siguientes ecuaciones para evaluar los coeficientes:

$$b_0 = f(x_0) \tag{18.8}$$

$$b_1 = f[x_1, x_0] \tag{18.9}$$

$$b_2 = f[x_2, x_1, x_0] \tag{18.10}$$

$$\vdots$$

$$b_n = f[x_n, x_{n-1}, \cdots, x_1, x_0] \tag{18.11}$$

donde las evaluaciones de la función colocadas entre paréntesis son diferencias divididas finitas. Por ejemplo, la primera diferencia dividida finita en forma general se representa como

$$f[x_i, x_j] = \frac{f(x_i) - f(x_j)}{x_i - x_j} \tag{18.12}$$

La *segunda diferencia dividida finita*, que representa la diferencia de las dos primeras diferencias divididas, se expresa en forma general como

$$f[x_i, x_j, x_k] = \frac{f[x_i, x_j] - f[x_j, x_k]}{x_i - x_k} \tag{18.13}$$

En forma similar, la *n-ésima diferencia dividida finita* es

$$f[x_n, x_{n-1}, \ldots, x_1, x_0] = \frac{f[x_n, x_{n-1}, \ldots, x_1] - f[x_{n-1}, x_{n-2}, \ldots, x_0]}{x_n - x_0} \quad (18.14)$$

Estas diferencias sirven para evaluar los coeficientes en las ecuaciones (18.8) a (18.11), los cuales se sustituirán en la ecuación (18.7) para obtener el polinomio de interpolación

$$f_n(x) = f(x_0) + (x - x_0)\, f[x_1, x_0] + (x - x_0)(x - x_1)\, f[x_2, x_1, x_0] + \cdots + (x - x_0)(x - x_1)\cdots(x - x_{n-1})\, f[x_n, x_{n-1}, \cdots, x_0] \quad (18.15)$$

que se conoce como *polinomio de interpolación de Newton en diferencias divididas*. Debe observarse que no se requiere que los puntos asociados con datos utilizados en la ecuación (18.15) estén igualmente espaciados o que los valores de la abscisa estén en orden ascendente, como se ilustra en el siguiente ejemplo. También, advierta cómo las ecuaciones (18.12) a (18.14) son recursivas, es decir, las diferencias de grado superior se calculan tomando diferencias de grado inferior (figura 18.5). Tal propiedad se aprovechará cuando desarrollemos un programa computacional eficiente en la sección 18.1.5 para implementar el método.

i	x_i	$f(x_i)$	Primero	Segundo	Tercero
0	x_0	$f(x_0)$	$f[x_1, x_0]$	$f[x_2, x_1, x_0]$	$f[x_3, x_2, x_1, x_0]$
1	x_1	$f(x_1)$	$f[x_2, x_1]$	$f[x_3, x_2, x_1]$	
2	x_2	$f(x_2)$	$f[x_3, x_2]$		
3	x_3	$f(x_3)$			

Figura 18.5
Representación gráfica de la naturaleza recursiva de las diferencias divididas finitas.

EJEMPLO 18.3 Polinomios de interpolación de Newton en diferencias divididas

Planteamiento del problema En el ejemplo 18.2, los puntos asociados con datos $x_0 = 1$, $x_1 = 4$ y $x_2 = 6$ se utilizaron para estimar ln 2 mediante una parábola. Ahora, agregando un cuarto punto [$x_3 = 5$; $f(x_3) = 1.609438$], estime ln 2 con un polinomio de interpolación de Newton de tercer grado.

Solución Utilizando la ecuación (18.7), con $n = 3$, el polinomio de tercer grado es

$$f_3(x) = b_0 + b_1(x - x_0) + b_2(x - x_0)(x - x_1) + b_3(x - x_0)(x - x_1)(x - x_2)$$

Las primeras diferencias divididas del problema son [ecuación (18.12)]

$$f[x_1, x_0] = \frac{1.386294 - 0}{4 - 0} = 0.4620981$$

$$f[x_2, x_1] = \frac{1.791759 - 1.386294}{6 - 4} = 0.2027326$$

$$f[x_3, x_2] = \frac{1.609438 - 1.791759}{5 - 6} = 0.1823216$$

Las segundas diferencias divididas son [ecuación (18.13)]

$$f[x_2, x_1, x_0] = \frac{0.2027326 - 0.4620981}{6 - 1} = -0.05187311$$

$$f[x_3, x_2, x_1] = \frac{0.1823216 - 0.2027326}{5 - 4} = -0.02041100$$

La tercera diferencia dividida es [ecuación (18.14) con $n = 3$]

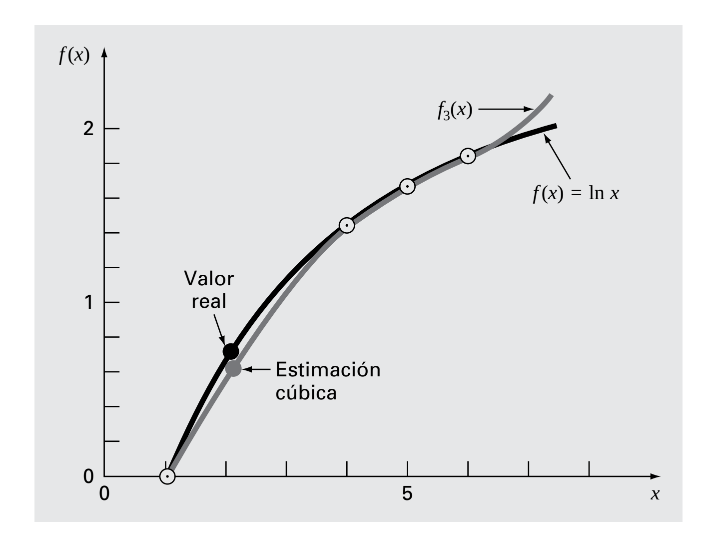


Figura 18.6
Uso de la interpolación cúbica para estimar ln 2.

$$f[x_3, x_2, x_1, x_0] = \frac{-0.02041100 - (-0.05187311)}{5-1} = 0.007865529$$

Los resultados de $f[x_1, x_0]$, $f[x_2, x_1, x_0]$ y $f[x_3, x_2, x_1, x_0]$ representan los coeficientes b_1, b_2 y b_3 de la ecuación (18.7), respectivamente. Junto con $b_0 = f(x_0) = 0.0$, la ecuación (18.7) es

$$f_3(x) = 0 + 0.4620981(x-1) - 0.05187311(x-1)(x-4) + 0.007865529(x-1)(x-4)(x-6)$$

la cual sirve para evaluar $f_3(2) = 0.6287686$, que representa un error relativo: $\varepsilon_t = 9.3\%$. La gráfica del polinomio cúbico se muestra en la figura 18.6.

18.1.4 Errores de la interpolación polinomial de Newton

Observe que la estructura de la ecuación (18.15) es similar a la expansión de la serie de Taylor en el sentido de que se van agregando términos en forma secuencial, para mostrar el comportamiento de grado superior de la función. Estos términos son diferencias divididas finitas y, así, representan aproximaciones de las derivadas de grado superior. En consecuencia, como ocurrió con la serie de Taylor, si la función verdadera es un polinomio de n-ésimo grado, entonces el polinomio de interpolación de n-ésimo grado basado en $n + 1$ puntos asociados con datos dará resultados exactos.

También, como en el caso de la serie de Taylor, es posible obtener una formulación para el error de truncamiento. De la ecuación (4.6) recuerde que el error de truncamiento en la serie de Taylor se expresa en forma general como

$$R_n = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x_{i+1} - x_i)^{n+1} \tag{4.6}$$

donde ξ está en alguna parte del intervalo de x_i a x_{i+1}. Para un polinomio de interpolación de n-ésimo grado, una expresión análoga para el error es

$$R_n = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - x_0)(x - x_1)\cdots(x - x_n) \tag{18.16}$$

donde ξ está en alguna parte del intervalo que contiene la incógnita y los datos. Para que esta fórmula sea útil, la función en turno debe ser conocida y diferenciable. Por lo común éste no es el caso. Por fortuna, hay una formulación alternativa que no requiere del conocimiento previo de la función. Utilizándose una diferencia dividida finita para aproximar la $(n + 1)$-ésima derivada,

$$R_n = f[x, x_n, x_{n-1}, \ldots, x_0](x - x_0)(x - x_1)\cdots(x - x_n) \tag{18.17}$$

donde $f[x, x_n, x_{n-1}, \ldots, x_0]$ es la $(n + 1)$-ésima diferencia dividida finita. Debido a que la ecuación (18.17) contiene la incógnita $f(x)$, no permite obtener el error. Sin embargo, si se tiene un punto asociado con un dato más, $f(x_{n+1})$, la ecuación (18.17) puede usarse para estimar el error como sigue:

$$R_n \cong f[x_{n+1}, x_n, x_{n-1}, \ldots, x_0](x - x_0)(x - x_1) \cdots (x - x_n) \tag{18.18}$$

EJEMPLO 18.4 Estimación del error para el polinomio de Newton

Planteamiento del problema Con la ecuación (18.18) estime el error en la interpolación polinomial de segundo grado del ejemplo 18.2. Use el punto asociado con un dato adicional $f(x_3) = f(5) = 1.609438$ para obtener sus resultados.

Solución Recuerde que en el ejemplo 18.2 el polinomio de interpolación de segundo grado proporcionó una estimación, $f_2(2) = 0.5658444$, que representa un error de 0.6931472 – 0.5658444 = 0.1273028. Si no se hubiera conocido el valor verdadero, como usualmente sucede, la ecuación (18.18), junto con el valor adicional en x_3, pudo haberse utilizado para estimar el error,

$$R_2 = f[x_3, x_2, x_1, x_0](x - x_0)(x - x_1)(x - x_2)$$

o

$$R_2 = 0.007865529(x - 1)(x - 4)(x - 6)$$

donde el valor de la diferencia dividida finita de tercer grado es como se calculó antes en el ejemplo 18.3. Esta expresión se evalúa en $x = 2$ para obtener

$$R_2 = 0.007865529(2 - 1)(2 - 4)(2 - 6) = 0.0629242$$

que es del mismo grado de magnitud que el error verdadero.

Con el ejemplo anterior y la ecuación (18.18), debe resultar claro que el error estimado para el polinomio de n-ésimo grado es equivalente a la diferencia entre las predicciones de grado $(n + 1)$ y de grado n. Es decir,

$$R_n = f_{n+1}(x) - f_n(x) \tag{18.19}$$

En otras palabras, el incremento que se agrega al caso de grado n para crear el caso de grado $(n + 1)$ [es decir, la ecuación (18.18)] se interpreta como un estimado del error de grado n. Esto se percibe con claridad al reordenar la ecuación (18.19):

$$f_{n+1}(x) = f_n(x) + R_n$$

La validez de tal procedimiento se refuerza por el hecho de que la serie es altamente convergente. En tal situación, la predicción del grado $(n + 1)$ debería ser mucho más cercana al valor verdadero que la predicción de grado n. En consecuencia, la ecuación (18.19) concuerda con nuestra definición estándar de error, al representar la diferencia entre la verdad y una aproximación. No obstante, observe que mientras todos los otros errores estimados para los procedimientos iterativos presentados hasta ahora se encontraron como una predicción presente menos una previa, la ecuación (18.19) constituye una predicción futura menos una presente. Lo anterior significa que para una serie que es de convergencia rápida, el error estimado de la ecuación (18.19) podría ser menor que el error verdadero. Esto representaría una calidad muy poco atractiva si el error estimado fuera a emplearse como un criterio de terminación. Sin embargo, como se expondrá en la siguiente sección, los polinomios de interpolación de grado superior son muy sensibles a errores en los datos (es decir, están mal condicionados). Cuando se emplean para interpolación, a menudo dan predicciones que divergen en forma significativa del valor verdadero. Si se trata de detectar errores, la ecuación (18.19) es más sensible a tal divergencia. De esta manera, es más valiosa con la clase de análisis de datos exploratorios para los que el polinomio de Newton es el más adecuado.

18.1.5 Algoritmo computacional para el polinomio de interpolación de Newton

Tres propiedades hacen a los polinomios de interpolación de Newton muy atractivos para aplicaciones en computadora:

1. Como en la ecuación (18.7), es posible desarrollar de manera secuencial versiones de grado superior con la adición de un solo término a la ecuación de grado inferior. Esto facilita la evaluación de algunas versiones de diferente grado en el mismo programa. En especial tal capacidad es valiosa cuando el grado del polinomio no se conoce *a priori*. Al agregar nuevos términos en forma secuencial, podemos determinar cuándo se alcanza un punto de regreso disminuido (es decir, cuando la adición de términos de grado superior ya no mejora de manera significativa la estimación, o en ciertas situaciones incluso la aleja). Las ecuaciones para estimar el error, que se analizan en el punto 3, resultan útiles para visualizar un criterio objetivo para identificar este punto de términos disminuidos.
2. Las diferencias divididas finitas que constituyen los coeficientes del polinomio [ecuaciones (18.8) hasta (18.11)] se pueden calcular eficientemente. Es decir, como en la ecuación (18.14) y la figura 18.5, las diferencias de grado inferior sirven para calcular las diferencias de grado mayor. Utilizando esta información previamente determinada, los coeficientes se calculan de manera eficiente. El algoritmo en la figura 18.7 incluye un esquema así.
3. El error estimado [ecuación (18.18)] se incorpora con facilidad en un algoritmo computacional debido a la manera secuencial en la cual se construye la predicción.

```
SUBROUTINE NewtInt (x, y, n, xi, yint, ea)
  LOCAL fdd(n,n)
  DOFOR i = 0, n
    fdd(i,0) = y(i)
  END DO
  DOFOR j = 1, n
    DOFOR i = 0, n - j
      fdd(i,j) = (fdd(i+1,j-1) - fdd(i,j-1))/(x(i+j) - x(i))
    END DO
  END DO
  xterm = 1
  yint(0) = fdd(0,0)
  DOFOR order = 1, n
    xterm = xterm * (xi - x(order-1))
    yint2 = yint(order-1) + fdd(0,order) * xterm
    ea(order-1) = yint2 - yint(order-1)
    yint(order) = yint2
  END order
END NewtInt
```

Figura 18.7
Un algoritmo para el polinomio de interpolación de Newton escrito en pseudocódigo.

Todas las características anteriores pueden aprovecharse e incorporarse en un algoritmo general para implementar el polinomio de Newton (figura 18.7). Observe que el algoritmo consiste de dos partes: la primera determina los coeficientes a partir de la ecuación (18.7); la segunda establece las predicciones y sus errores correspondientes. La utilidad de dicho algoritmo se demuestra en el siguiente ejemplo.

EJEMPLO 18.5 Estimaciones del error para determinar el grado de interpolación adecuado

Planteamiento del problema Después de incorporar el error [ecuación (18.18)], utilice el algoritmo computacional que se muestra en la figura 18.7 y la información tabulada siguiente para evaluar $f(x) = \ln x$ en $x = 2$:

x	$f(x) = \ln x$
1	0
4	1.3862944
6	1.7917595
5	1.6094379
3	1.0986123
1.5	0.4054641
2.5	0.9162907
3.5	1.2527630

Solución Los resultados de emplear el algoritmo de la figura 18.7 para obtener una solución se muestran en la figura 18.8. El error estimado, junto con el error verdadero (basándose en el hecho de que ln 2 = 0.6931472), se ilustran en la figura 18.9. Observe que el error estimado y el error verdadero son similares y que su concordancia mejora conforme aumenta el grado. A partir de estos resultados se concluye que la versión de quinto grado da una buena estimación y que los términos de grado superior no mejoran significativamente la predicción.

Este ejercicio también ilustra la importancia de la posición y el orden de los puntos. Por ejemplo, hasta la estimación de tercer grado, la mejoría es lenta debido a que los puntos que se agregaron (en $x = 4$, 6 y 5) están distantes y a un lado del punto de análisis en $x = 2$. La estimación de cuarto grado muestra una mejoría un poco mayor, ya

```
NUMERO DE PUNTOS? 8
X( 0 ), y( 0 ) = ? 1,0
X( 1 ), y( 1 ) = ? 4,1.3862944
X( 2 ), y( 2 ) = ? 6,1.7917595
X( 3 ), y( 3 ) = ? 5,1.6094379
X( 4 ), y( 4 ) = ? 3,1.0986123
X( 5 ), y( 5 ) = ? 1.5,0.40546411
X( 6 ), y( 6 ) = ? 2.5,0.91629073
X( 7 ), y( 7 ) = ? 3.5,1.2527630

INTERPOLACION EN X = 2
GRADO   F(X)           ERROR
0       0.000000       0.462098
1       0.462098       0.103746
2       0.565844       0.062924
3       0.628769       0.046953
4       0.675722       0.021792
5       0.697514       -0.003616
6       0.693898       -0.000459
7       0.693439
```

Figura 18.8
Resultados de un programa, basado en el algoritmo de la figura 18.7, para evaluar ln 2.

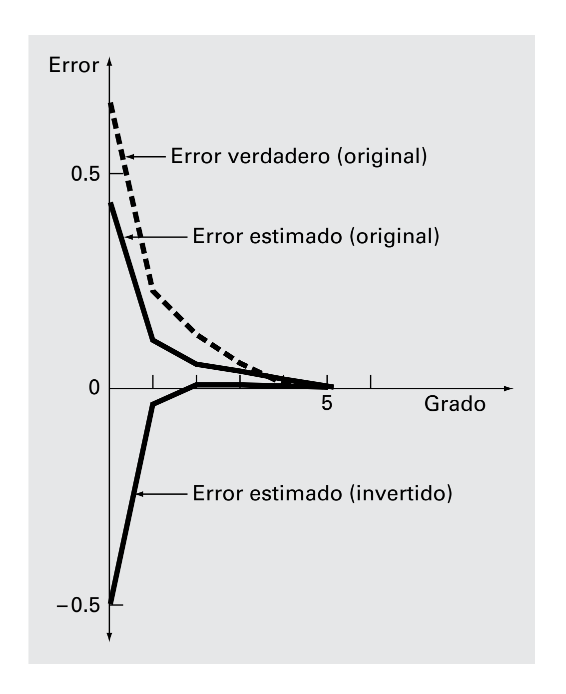


Figura 18.9
Errores relativos porcentuales para la predicción de ln 2 como función del grado del polinomio de interpolación.

que el nuevo punto en $x = 3$ está más cerca de la incógnita. Aunque, la disminución más dramática en el error corresponde a la inclusión del término de quinto grado usando el punto asociado con un dato en $x = 1.5$. Dicho punto está cerca de la incógnita y también se halla al lado opuesto de la mayoría de los otros puntos. En consecuencia, el error se reduce a casi un grado de magnitud.

La importancia de la posición y el orden de los datos también se demuestra al usar los mismos datos para obtener una estimación para ln 2, pero considerando los puntos en un orden diferente. La figura 18.9 muestra los resultados en el caso de invertir el orden de los datos originales; es decir, $x_0 = 3.5$, $x_1 = 2.5$, $x_3 = 1.5$, y así sucesivamente. Como los puntos iniciales en este caso se hallan más cercanos y espaciados a ambos lados de ln 2, el error disminuye mucho más rápidamente que en la situación original. En el término de segundo grado, el error se redujo a menos de $\varepsilon_t = 2\%$. Se podrían emplear otras combinaciones para obtener diferentes velocidades de convergencia.

El ejemplo anterior ilustra la importancia de la selección de los puntos. Como es intuitivamente lógico, los puntos deberían estar centrados alrededor, y tan cerca como sea posible, de las incógnitas. Esta observación también se sustenta por un análisis directo de la ecuación para estimar el error [ecuación (18.17)]. Si suponemos que las diferencias divididas finitas no varían mucho a través de los datos, el error es proporcional al producto: $(x - x_0)(x - x_1) \cdots (x - x_n)$. Obviamente, cuanto más cercanos a x estén los puntos, menor será la magnitud de este producto.

18.2 POLINOMIOS DE INTERPOLACIÓN DE LAGRANGE

El *polinomio de interpolación de Lagrange* es simplemente una reformulación del polinomio de Newton que evita el cálculo de las diferencias divididas, y se representa de manera concisa como

$$f_n(x) = \sum_{i=0}^{n} L_i(x) f(x_i) \tag{18.20}$$

donde

$$L_i(x) = \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{x - x_j}{x_i - x_j} \tag{18.21}$$

donde Π designa el "producto de". Por ejemplo, la versión lineal (n = 1) es

$$f_1(x) = \frac{x - x_1}{x_0 - x_1} f(x_0) + \frac{x - x_0}{x_1 - x_0} f(x_1) \tag{18.22}$$

y la versión de segundo grado es

$$f_2(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} f(x_0) + \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} f(x_1) + \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} f(x_2) \tag{18.23}$$

La ecuación (18.20) se obtiene de manera directa del polinomio de Newton (cuadro 18.1). Sin embargo, el razonamiento detrás de la formulación de Lagrange se comprende directamente al darse cuenta de que cada término $L_i(x)$ será 1 en $x = x_i$ y 0 en todos los otros puntos (figura 18.10). De esta forma, cada producto $L_i(x)\, f(x_i)$ toma el valor de $f(x_i)$ en el punto x_i. En consecuencia, la sumatoria de todos los productos en la ecuación (18.20) es el único polinomio de n-ésimo grado que pasa exactamente a través de todos los n + 1 puntos asociados con datos.

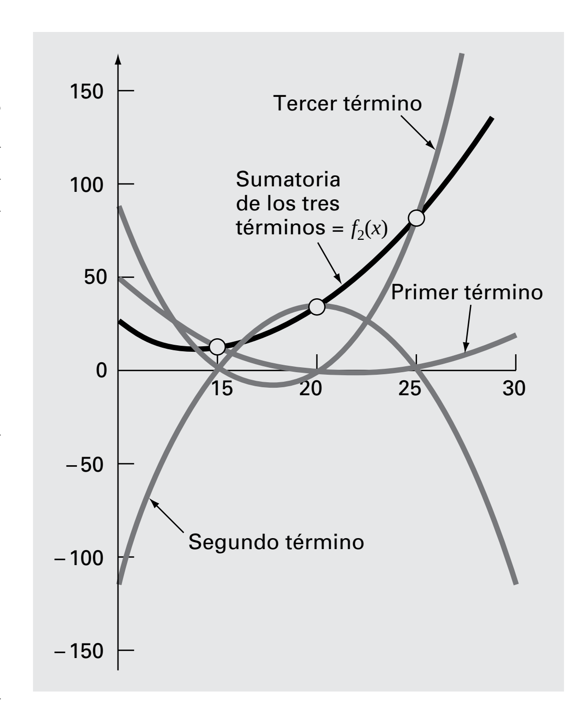


Figura 18.10
Descripción visual del razonamiento detrás del polinomio de Lagrange. Esta figura muestra un caso de segundo grado. Cada uno de los tres términos en la ecuación (18.23) pasa a través de uno de los puntos asociados con datos y es cero en los otros dos. La suma de los tres términos, por lo tanto, debe ser el único polinomio de segundo grado $f_2(x)$ que pasa exactamente a través de los tres puntos.

EJEMPLO 18.6 Polinomios de interpolación de Lagrange

Planteamiento del problema Con un polinomio de interpolación de Lagrange de primero y segundo grado evalúe ln 2 basándose en los datos del ejemplo 18.2:

$x_0 = 1$ $f(x_0) = 0$

$x_1 = 4$ $f(x_1) = 1.386294$

$x_2 = 6$ $f(x_2) = 1.791760$

Solución El polinomio de primer grado [ecuación (18.22)] se utiliza para obtener la estimación en x = 2,

$$f_1(2) = \frac{2-4}{1-4} 0 + \frac{2-1}{4-1} 1.386294 = 0.4620981$$

De manera similar, el polinomio de segundo grado se desarrolla así: [ecuación (18.23)]

$$f_2(2) = \frac{(2-4)(2-6)}{(1-4)(1-6)} 0 + \frac{(2-1)(2-6)}{(4-1)(4-6)} 1.386294 + \frac{(2-1)(2-4)}{(6-1)(6-4)} 1.791760 = 0.5658444$$

Como se esperaba, ambos resultados concuerdan con los que se obtuvieron antes al usar el polinomio de interpolación de Newton.



Observe que, como en el método de Newton, la forma de Lagrange tiene un error estimado de [ecuación (18.17)]

$$R_n = f[x, x_n, x_{n-1}, \ldots, x_0] \prod_{i=0}^{n} (x - x_i)$$

De este modo, si se tiene un punto adicional en $x = x_{n+1}$, se puede obtener un error estimado. Sin embargo, como no se emplean las diferencias divididas finitas como parte del algoritmo de Lagrange, esto se hace rara vez.

Las ecuaciones (18.20) y (18.21) se programan de manera muy simple para implementarse en una computadora. La figura 18.11 muestra el pseudocódigo que sirve para tal propósito.

En resumen, en los casos donde se desconoce el grado del polinomio, el método de Newton tiene ventajas debido a la comprensión que proporciona respecto al comportamiento de las fórmulas de diferente grado. Además, el estimado del error representado por la ecuación (18.18) se agrega usualmente en el cálculo del polinomio de Newton debido a que el estimado emplea una diferencia finita (ejemplo 18.5). De esta manera, para cálculos exploratorios, a menudo se prefiere el método de Newton.

Cuadro 18.1 Obtención del polinomio de lagrange directamente a partir del polinomio de interpolación de Newton

El polinomio de interpolación de Lagrange se obtiene de manera directa a partir de la formulación del polinomio de Newton. Haremos esto únicamente en el caso del polinomio de primer grado [ecuación (18.2)]. Para obtener la forma de Lagrange, reformulamos las diferencias divididas. Por ejemplo, la primera diferencia dividida,

$$f[x_1, x_0] = \frac{f(x_1) - f(x_0)}{x_1 - x_0} \tag{C18.1.1}$$

se reformula como

$$f[x_1, x_0] = \frac{f(x_1)}{x_1 - x_0} + \frac{f(x_0)}{x_0 - x_1} \tag{C18.1.2}$$

conocida como la *forma simétrica.* Al sustituir la ecuación (C18.1.2) en la (18.2) se obtiene

$$f_1(x) = f(x_0) + \frac{x - x_0}{x_1 - x_0} f(x_1) + \frac{x - x_0}{x_0 - x_1} f(x_0)$$

Por último, al agrupar términos semejantes y simplificar se obtiene la forma del polinomio de Lagrange,

$$f_1(x) = \frac{x - x_1}{x_0 - x_1} f(x_0) + \frac{x - x_0}{x_1 - x_0} f(x_1)$$

```
FUNCTION Lagrng(x, y, n, xx)
  sum = 0
  DOFOR i = 0, n
    product = y_i
    DOFOR j = 0, n
      IF i ≠ j THEN
         product = product*(xx - x_j)/(x_i - x_j)
      ENDIF
    END DO
    sum = sum + product
  END DO
  Lagrng = sum
END Lagrng
```

Figura 18.11
Pseudocódigo para la interpolación de Lagrange. Este algoritmo se establece para calcular una sola predicción de grado *n*-ésimo, donde $n + 1$ es el número de puntos asociados con datos.

Cuando se va a ejecutar sólo una interpolación, las formulaciones de Lagrange y de Newton requieren un trabajo computacional semejante. No obstante, la versión de Lagrange es un poco más fácil de programar. Debido a que no requiere del cálculo ni del almacenaje de diferencias divididas, la forma de Lagrange a menudo se utiliza cuando el grado del polinomio se conoce *a priori*.

EJEMPLO 18.7 Interpolación de Lagrange empleando la computadora

Planteamiento del problema Es posible usar el algoritmo de la figura 18.11 para estudiar un problema de análisis de tendencia que se relaciona con nuestro conocido caso de la caída del paracaidista. Suponga que se tiene un instrumento para medir la velocidad del paracaidista. Los datos obtenidos en una prueba particular se encuentran en la tabla de la derecha.

Tiempo, s	Velocidad medida *v*, cm/s
1	800
3	2 310
5	3 090
7	3 940
13	4 755

Nuestro problema consiste en estimar la velocidad del paracaidista en $t = 10$ s para tener las mediciones faltantes entre $t = 7$ y $t = 13$ s. Estamos conscientes de que el comportamiento de los polinomios de interpolación tal vez resulte inesperado. Por lo tanto, construiremos polinomios de grados 4, 3, 2 y 1, y compararemos los resultados.

Solución El algoritmo de Lagrange se utiliza para construir polinomios de interpolación de cuarto, tercer, segundo y primer grados.

El polinomio de cuarto grado y los datos de entrada se grafican como se muestra en la figura 18.12*a*. Es evidente, al observar la gráfica, que el valor estimado de y en $x = 10$ es mayor que la tendencia global de los datos.

Las figuras 18.12*b* a 18.12*d* muestran las gráficas de los resultados de los cálculos con las interpolaciones de los polinomios de tercer, segundo y primer grados, respectivamente. Se observa que cuanto más bajo sea el grado, menor será el valor estimado de la velocidad en $t = 10$ s. Las gráficas de los polinomios de interpolación indican que los polinomios de grado superior tienden a sobrepasar la tendencia de los datos, lo cual sugiere que las versiones de primer o segundo grado son las más adecuadas para este análisis de tendencia en particular. No obstante, debe recordarse que debido a que tratamos con datos inciertos, la regresión, de hecho, será la más adecuada.

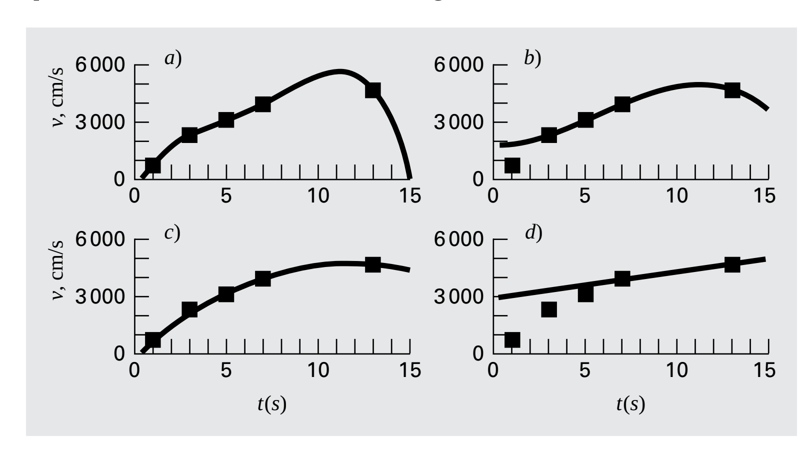


Figura 18.12
Gráficas que muestran interpolaciones de *a*) cuarto grado, *b*) tercer grado, *c*) segundo grado y *d*) primer grado.

El ejemplo anterior ilustró que los polinomios de grado superior tienden a estar mal condicionados; es decir, tienden a ser altamente susceptibles a los errores de redondeo. El mismo problema se presenta en la regresión con polinomios de grado superior. La aritmética de doble precisión ayuda algunas veces a disminuir el problema. Sin embargo, conforme el grado aumente, habrá un punto donde el error de redondeo interferirá con la habilidad para interpolar usando los procedimientos simples estudiados hasta ahora.

18.3 COEFICIENTES DE UN POLINOMIO DE INTERPOLACIÓN

Aunque el polinomio de Newton y el de Lagrange son adecuados para determinar valores intermedios entre puntos, no ofrecen un polinomio adecuado de la forma convencional

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \tag{18.24}$$

Un método directo para calcular los coeficientes de este polinomio se basa en el hecho de que se requieren $n + 1$ puntos asociados con datos para determinar los $n + 1$ coeficientes. Así, se utiliza un sistema de ecuaciones algebraicas lineales simultáneas para calcular las a_i. Por ejemplo, suponga que usted desea calcular los coeficientes de la parábola

$$f(x) = a_0 + a_1x + a_2x^2 \tag{18.25}$$

Se requiere de tres puntos: $[x_0, f(x_0)]$, $[x_1, f(x_1)]$ y $[x_2, f(x_2)]$. Cada uno se sustituye en la ecuación (18.25):

$$\begin{aligned} f(x_0) &= a_0 + a_1x_0 + a_2x_0^2 \\ f(x_1) &= a_0 + a_1x_1 + a_2x_1^2 \\ f(x_2) &= a_0 + a_1x_2 + a_2x_2^2 \end{aligned} \tag{18.26}$$

De esta manera, las x son los puntos conocidos, y las a_i las incógnitas. Como hay el mismo número de ecuaciones que de incógnitas, la ecuación (18.26) se podría resolver con uno de los métodos de eliminación de la parte tres.

Debe observarse que el procedimiento anterior no es el método de interpolación más eficiente para determinar los coeficientes de un polinomio. Press *et al*. (1992) ofrecen un análisis y códigos para computadora de los procedimientos más eficientes. Cualquiera que sea la técnica empleada, se debe hacer una advertencia. Sistemas como los de la ecuación (18.26) están notoriamente mal condicionados. Ya sea que se resuelvan con un método de eliminación o con un algoritmo más eficiente, los coeficientes resultantes pueden ser bastante inexactos, en particular para n grandes. Si se usan para una interpolación subsecuente, a menudo dan resultados erróneos.

En resumen, si usted se interesa en determinar un punto intermedio, emplee la interpolación de Newton o de Lagrange. Si tiene que determinar una ecuación de la forma de la (18.24), limítese a polinomios de grado menor y verifique cuidadosamente sus resultados.

18.4 INTERPOLACIÓN INVERSA

Como la nomenclatura implica, los valores de $f(x)$ y x en la mayoría de los problemas de interpolación son las variables dependiente e independiente, respectivamente. En consecuencia, los valores de las x con frecuencia están espaciados uniformemente. Un ejemplo simple es una tabla de valores obtenida para la función $f(x) = 1/x$,

x	1	2	3	4	5	6	7
$f(x)$	1	0.5	0.3333	0.25	0.2	0.1667	0.1429

Ahora suponga que usted debe usar los mismos datos, pero que se le ha dado un valor de $f(x)$ y debe determinar el valor correspondiente de x. Por ejemplo, para los datos anteriores, suponga que se le pide determinar el valor de x que corresponda a $f(x) = 0.3$. En tal caso, como se tiene la función y es fácil de manipular, la respuesta correcta se determina directamente, $x = 1/0.3 = 3.3333$.

A ese problema se le conoce como *interpolación inversa*. En un caso más complicado, usted puede sentirse tentado a intercambiar los valores $f(x)$ y x [es decir, tan sólo graficar x contra $f(x)$] y usar un procedimiento como la interpolación de Lagrange para determinar el resultado. Por desgracia, cuando usted invierte las variables no hay garantía de que los valores junto con la nueva abscisa [las $f(x)$] estén espaciados de una manera uniforme. Es más, en muchos casos, los valores estarán "condensados". Es decir, tendrán la apariencia de una escala logarítmica, con algunos puntos adyacentes muy amontonados y otros muy dispersos. Por ejemplo, para $f(x) = 1/x$ el resultado es

$f(x)$	0.1429	0.1667	0.2	0.25	0.3333	0.5	1
x	7	6	5	4	3	2	1

Tal espaciamiento no uniforme en las abscisas a menudo lleva a oscilaciones en el resultado del polinomio de interpolación. Esto puede ocurrir aun para polinomios de grado inferior.

Una estrategia alterna es ajustar un polinomio de interpolación de grado *n*-ésimo, $f_n(x)$, a los datos originales [es decir, con $f(x)$ contra x]. En la mayoría de los casos, como las x están espaciadas de manera uniforme, este polinomio no estará mal condicionado. La respuesta a su problema, entonces, consiste en encontrar el valor de x que haga este polinomio igual al dado por $f(x)$. Así, ¡el problema de interpolación se reduce a un problema de raíces!

Por ejemplo, para el problema antes descrito, un procedimiento simple sería ajustar los tres puntos a un polinomio cuadrático: (2, 0.5), (3, 0.3333) y (4, 0.25), cuyo resultado será

$$f_2(x) = 1.08333 - 0.375x + 0.041667x^2$$

La respuesta al problema de interpolación inversa para determinar la x correspondiente a $f(x) = 0.3$ será equivalente a la determinación de las raíces de

$$0.3 = 1.08333 - 0.375x + 0.041667x^2$$

En este caso simple, la fórmula cuadrática se utiliza para calcular

$$x = \frac{0.375 \pm \sqrt{(-0.375)^2 - 4(0.041667)0.78333}}{2(0.041667)} = \frac{5.704158}{3.295842}$$

Así, la segunda raíz, 3.296, es una buena aproximación al valor verdadero: 3.333. Si se desea una exactitud adicional, entonces podría emplear un polinomio de tercero o cuarto grados junto con uno de los métodos para la localización de raíces analizado en la parte dos.

18.5 COMENTARIOS ADICIONALES

Antes de proceder con la siguiente sección, se deben mencionar dos temas adicionales: la interpolación y extrapolación con datos igualmente espaciados.

Como ambos polinomios, el de Newton y el de Lagrange, son compatibles con datos espaciados en forma arbitraria, usted se preguntará por qué nos ocupamos del caso especial de datos igualmente espaciados (cuadro 18.2). Antes de la llegada de las computadoras digitales, dichas técnicas tenían gran utilidad para interpolación a partir de tablas con datos igualmente espaciados. De hecho, se desarrolló una estructura computacional, conocida como tabla de diferencias divididas, para facilitar la implementación de dichas técnicas. (La figura 18.5 es un ejemplo de esa tabla.)

Sin embargo, como las fórmulas dadas son subconjuntos de los esquemas de Newton y Lagrange compatibles con una computadora y debido a las muchas funciones tabulares existentes, como subrutinas de bibliotecas, ha disminuido la necesidad de tener versiones para datos igualmente espaciados. A pesar de ello, las hemos incluido en este tema por su relevancia en las últimas partes de este libro. En especial, son necesarias para obtener fórmulas de integración numérica que por lo común utilizan datos igualmente espaciados (capítulo 21). Como las fórmulas de integración numérica son importantes en la solución de ecuaciones diferenciales ordinarias, el análisis del cuadro 18.2 adquiere también significado para la parte siete.

Extrapolación es el proceso de estimar un valor de $f(x)$ que se encuentra fuera del dominio de los valores conocidos, $x_0, x_1, ..., x_n$ (figura 18.13). En una sección anterior, mencionamos que la interpolación más exacta se obtiene cuando las incógnitas están cerca de los puntos. En efecto, éste no es el caso cuando la incógnita se encuentra fuera del intervalo y, en consecuencia, el error en la extrapolación puede ser muy grande. Como se ilustra en la figura 18.13, la naturaleza de la extrapolación de extremos abiertos representa un paso a lo desconocido, ya que el proceso extiende la curva más allá de la región conocida. Como tal, la curva real podrá fácilmente diverger de la predicción. Por lo tanto, se debe tener mucho cuidado cuando aparezca un problema donde se deba extrapolar.

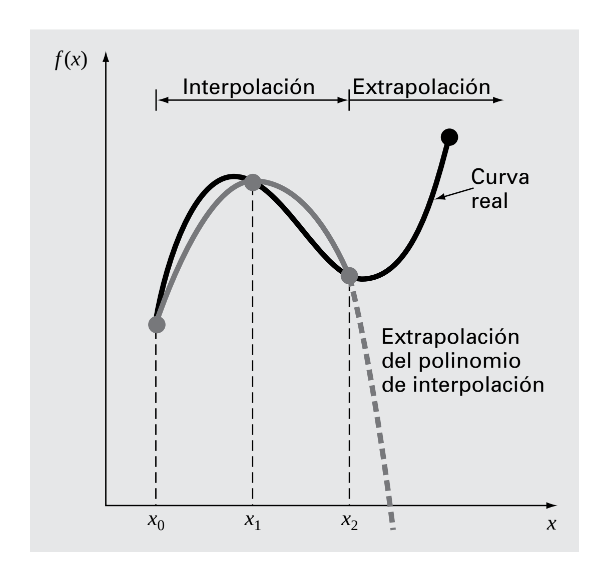


Figura 18.13
Ilustración de la posible divergencia de una predicción extrapolada. La extrapolación se basa en ajustar una parábola con los primeros tres puntos conocidos.

Cuadro 18.2 Interpolación con datos igualmente espaciados

Si los datos están igualmente espaciados y en orden ascendente, entonces la variable independiente tiene los valores de

$$x_1 = x_0 + h$$
$$x_2 = x_0 + 2h$$
$$\vdots$$
$$x_n = x_0 + nh$$

donde h es el intervalo, o tamaño de paso, entre los datos. Basándose en esto, las diferencias divididas finitas se pueden expresar en forma concisa. Por ejemplo, la segunda diferencia dividida hacia adelante es

$$f[x_0, x_1, x_2] = \frac{\dfrac{f(x_2) - f(x_1)}{x_2 - x_1} - \dfrac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0}$$

que se expresa como

$$f[x_0, x_1, x_2] = \frac{f(x_2) - 2f(x_1) + f(x_0)}{2h^2} \quad \text{(C18.2.1)}$$

ya que $x_1 - x_0 = x_2 - x_1 = (x_2 - x_0)/2 = h$. Ahora recuerde que la segunda diferencia hacia adelante es igual a [numerador de la ecuación (4.24)]

$$\Delta^2 f(x_0) = f(x_2) - 2f(x_1) + f(x_0)$$

Por lo tanto, la ecuación (C18.2.1) se representa como

$$f[x_0, x_1, x_2] = \frac{\Delta^2 f(x_0)}{2!h^2}$$

o, en general

$$f[x_0, x_1, \ldots, x_n] = \frac{\Delta^n f(x_0)}{n!h^n} \quad \text{(C18.2.2)}$$

Usando la ecuación (C18.2.2), en el caso de datos igualmente espaciados, expresamos el polinomio de interpolación de Newton [ecuación (18.15)] como

$$f_n(x) = f(x_0) + \frac{\Delta f(x_0)}{h}(x - x_0) + \frac{\Delta^2 f(x_0)}{2!h^2}(x - x_0)(x - x_0 - h) + \cdots + \frac{\Delta^n f(x_0)}{n!h^n}(x - x_0)(x - x_0 - h) \cdots [x - x_0 - (n-1)h] + R_n \quad \text{(C18.2.3)}$$

donde el residuo es el mismo que en la ecuación (18.16). Esta ecuación se conoce como *fórmula de Newton* o la *fórmula hacia adelante de Newton-Gregory*, que se puede simplificar más al definir una nueva cantidad, α:

$$\alpha = \frac{x - x_0}{h}$$

Esta definición se utiliza para desarrollar las siguientes expresiones simplificadas de los términos en la ecuación (C18.2.3):

$$x - x_0 = \alpha h$$
$$x - x_0 - h = \alpha h - h = h(\alpha - 1)$$
$$\vdots$$
$$x - x_0 - (n-1)h = \alpha h - (n-1)h = h(\alpha - n + 1)$$

que se sustituye en la ecuación (C18.2.3) para tener

$$f_n(x) = f(x_0) + \Delta f(x_0)\alpha + \frac{\Delta^2 f(x_0)}{2!}\alpha(\alpha - 1) + \cdots + \frac{\Delta^n f(x_0)}{n!}\alpha(\alpha - 1) \cdots (\alpha - n + 1) + R_n \quad \text{(C18.2.4)}$$

donde

$$R_n = \frac{f^{(n+1)}(\xi)}{(n+1)!}h^{n+1}\alpha(\alpha - 1)(\alpha - 2) \cdots (\alpha - n)$$

En el capítulo 21 esta notación concisa tendrá utilidad en la deducción y análisis del error de las fórmulas de integración.

Además de la fórmula hacia adelante, también existen las fórmulas hacia atrás y central de Newton-Gregory. Para más información respecto de la interpolación para datos igualmente espaciados véase Carnahan, Luther y Wilkes (1969).

18.6 INTERPOLACIÓN MEDIANTE TRAZADORES (SPLINES)

En las secciones anteriores se usaron polinomios de n-ésimo grado para interpolar entre $n + 1$ puntos asociados con datos. Por ejemplo, para ocho puntos se puede obtener un perfecto polinomio de séptimo grado. Esta curva podría agrupar todas las curvas (al menos hasta, e incluso, la séptima derivada) sugeridas por los puntos. No obstante, hay casos donde estas funciones llevarían a resultados erróneos a causa de los errores de redondeo y los puntos lejanos. Un procedimiento alternativo consiste en colocar polinomios de grado inferior en subconjuntos de los puntos asociados con datos. Tales polinomios conectores se denominan *trazadores* o *splines*.

Por ejemplo, las curvas de tercer grado empleadas para unir cada par de puntos asociados con datos se llaman *trazadores cúbicos*. Esas funciones se pueden construir de tal forma que las conexiones entre ecuaciones cúbicas adyacentes resulten visualmente suaves. Podría parecer que la aproximación de tercer grado de los trazadores sería inferior a la expresión de séptimo grado. Usted se preguntaría por qué un trazador aún resulta preferible.

La figura 18.14 ilustra una situación donde un trazador se comporta mejor que un polinomio de grado superior. Éste es el caso donde una función en general es suave, pero presenta un cambio abrupto en algún lugar de la región de interés. El tamaño de paso representado en la figura 18.14 es un ejemplo extremo de tal cambio y sirve para ilustrar esta idea.

La figura 18.14*a* a *c* ilustra cómo un polinomio de grado superior tiende a formar una curva de oscilaciones bruscas en la vecindad de un cambio súbito. En contraste, el trazador también une los puntos; pero como está limitado a cambios de tercer grado, las oscilaciones son mínimas. De esta manera, el trazador usualmente proporciona una mejor aproximación al comportamiento de las funciones que tienen cambios locales y abruptos.

El concepto de trazador se originó en la técnica de dibujo que usa una cinta delgada y flexible (llamada *spline*, en inglés), para dibujar curvas suaves a través de un conjunto de puntos. El proceso se representa en la figura 18.15 para una serie de cinco alfileres (puntos asociados con datos). En esta técnica, el dibujante coloca un papel sobre una mesa de madera y coloca alfileres o clavos en el papel (y la mesa) en la ubicación de los puntos asociados con datos. Una curva cúbica suave resulta al entrelazar la cinta entre los alfileres. De aquí que se haya adoptado el nombre de “trazador cúbico” (en inglés: “cubic spline”) para los polinomios de este tipo.

En esta sección se usarán primero funciones lineales simples para presentar algunos conceptos y problemas básicos relacionados con la interpolación mediante splines. A continuación obtendremos un algoritmo para el ajuste de trazadores cuadráticos a los datos. Por último, presentamos material sobre el trazador cúbico, que es la versión más común y útil en la práctica de la ingeniería.

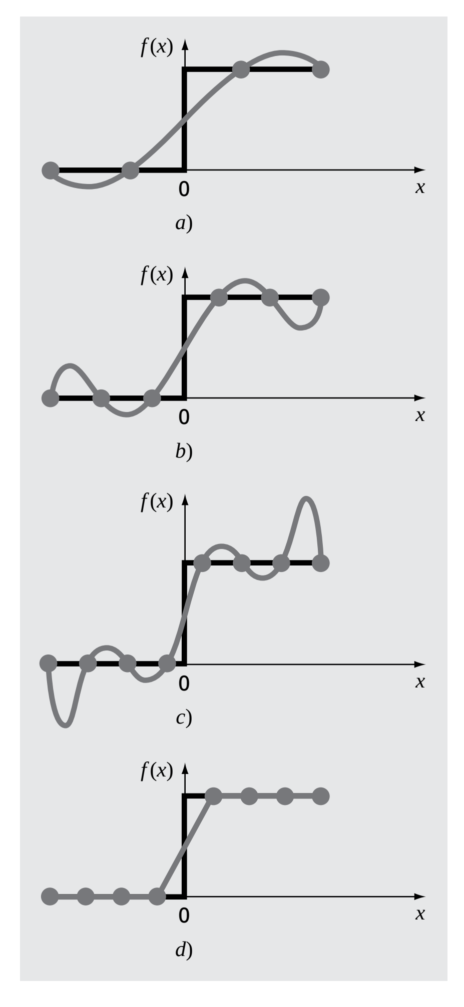


Figura 18.14
Una representación visual de una situación en la que los trazadores son mejores que los polinomios de interpolación de grado superior. La función que se ajusta presenta un incremento súbito en $x = 0$. Los incisos *a*) a *c*) indican que el cambio abrupto induce oscilaciones en los polinomios de interpolación. En contraste, como se limitan a curvas de tercer grado con transiciones suaves, un trazador lineal *d*) ofrece una aproximación mucho más aceptable.

18.6.1 Trazadores lineales

La unión más simple entre dos puntos es una línea recta. Los trazadores de primer grado para un grupo de puntos asociados con datos ordenados pueden definirse como un conjunto de funciones lineales,

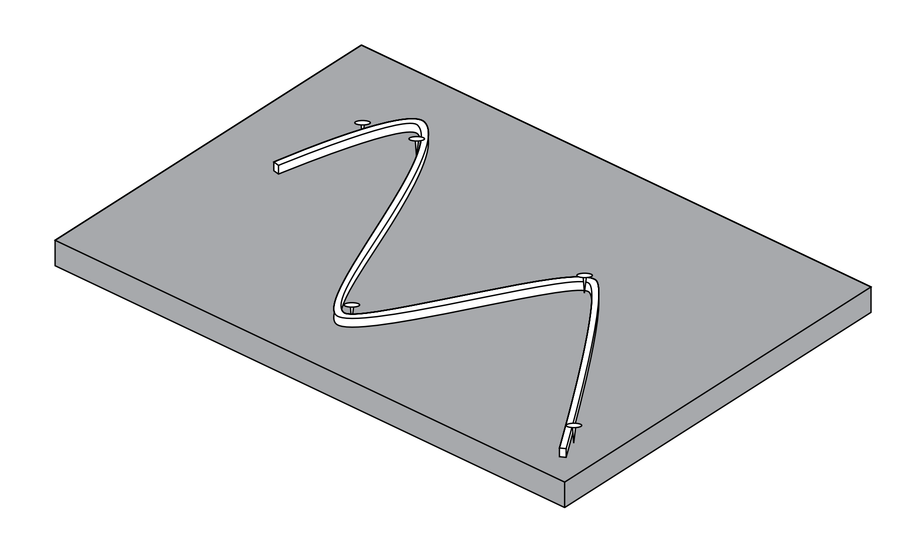


Figura 18.15
La técnica de dibujo que usa una cinta delgada y flexible para dibujar curvas suaves a través de una serie de puntos. Observe cómo en los puntos extremos, el trazador tiende a volverse recto. Esto se conoce como un trazador “natural”.

$$
\begin{aligned}
f(x) &= f(x_0) + m_0(x - x_0) && x_0 \le x \le x_1 \\
f(x) &= f(x_1) + m_1(x - x_1) && x_1 \le x \le x_2 \\
&\vdots \\
f(x) &= f(x_{n-1}) + m_{n-1}(x - x_{n-1}) && x_{n-1} \le x \le x_n
\end{aligned}
$$

donde m_i es la pendiente de la línea recta que une los puntos:

$$m_i = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} \tag{18.27}$$

Estas ecuaciones se pueden usar para evaluar la función en cualquier punto entre x_0 y x_n localizando primero el intervalo dentro del cual está el punto. Después se usa la ecuación adecuada para determinar el valor de la función dentro del intervalo. El método es obviamente idéntico al de la interpolación lineal.

EJEMPLO 18.8 Trazadores de primer grado

Tabla 18.1
Datos para ajustarse con trazadores.

x	$f(x)$
3.0	2.5
4.5	1.0
7.0	2.5
9.0	0.5

Planteamiento del problema Ajuste los datos de la tabla 18.1 con trazadores de primer grado. Evalúe la función en $x = 5$.

Solución Se utilizan los datos para determinar las pendientes entre los puntos. Por ejemplo, en el intervalo de $x = 4.5$ a $x = 7$ la pendiente se calcula con la ecuación (18.27):

$$m = \frac{2.5 - 1}{7 - 4.5} = 0.60$$

Se calculan las pendientes en los otros intervalos y los trazadores de primer grado obtenidos se grafican en la figura 18.16*a*. El valor en $x = 5$ es 1.3.

Una inspección visual a la figura 18.16*a* indica que la principal desventaja de los trazadores de primer grado es que no son suaves. En esencia, en los puntos asociados con datos donde se encuentran dos trazadores (llamado *nodo*), la pendiente cambia de forma abrupta. Formalmente, la primer derivada de la función es discontinua en esos puntos. Esta deficiencia se resuelve usando trazadores polinomiales de grado superior, que aseguren suavidad en los nodos al igualar las derivadas en esos puntos, como se analiza en la siguiente sección.

18.6.2 Trazadores (splines) cuadráticos

Para asegurar que las derivadas m-ésimas sean continuas en los nodos, se debe emplear un trazador de un grado de, al menos, $m + 1$. En la práctica se usan con más frecuencia polinomios de tercer grado o trazadores cúbicos que aseguran primera y segunda derivadas continuas. Aunque las derivadas de tercer grado y mayores podrían ser discontinuas cuando se usan trazadores cúbicos, por lo común no pueden detectarse en forma visual y, en consecuencia, se ignoran.

Debido a que la deducción de trazadores cúbicos es algo complicada, la hemos incluido en una sección subsecuente. Decidimos ilustrar primero el concepto de interpolación mediante trazadores usando polinomios de segundo grado. Esos "trazadores cuadráticos" tienen primeras derivadas continuas en los nodos. Aunque los trazadores cuadráticos no aseguran segundas derivadas iguales en los nodos, sirven muy bien para demostrar el procedimiento general en el desarrollo de trazadores de grado superior.

El objetivo de los trazadores cuadráticos es obtener un polinomio de segundo grado para cada intervalo entre los puntos asociados con datos. De manera general, el polinomio en cada intervalo se representa como

$$f_i(x) = a_i x^2 + b_i x + c_i \tag{18.28}$$

La figura 18.17 servirá para aclarar la notación. Para $n + 1$ puntos asociados con datos ($i = 0, 1, 2, ..., n$) existen n intervalos y, en consecuencia, $3n$ constantes desconocidas (las a, b y c) por evaluar. Por lo tanto, se requieren $3n$ ecuaciones o condiciones para evaluar las incógnitas. Éstas son:

Figura 18.16
Ajuste mediante trazadores de un conjunto de cuatro puntos. *a*) Trazador lineal, *b*) trazador cuadrático y *c*) trazador cúbico; se grafica también un polinomio de interpolación cúbico.

1. *Los valores de la función de polinomios adyacentes deben ser iguales en los nodos interiores.* Esta condición se representa como

$$a_{i-1}x_{i-1}^2 + b_{i-1}x_{i-1} + c_{i-1} = f(x_{i-1}) \tag{18.29}$$

$$a_i x_{i-1}^2 + b_i x_{i-1} + c_i = f(x_{i-1}) \tag{18.30}$$

para $i = 2$ a n. Como sólo se emplean nodos interiores, las ecuaciones (18.29) y (18.30) proporcionan, cada una, $n - 1$ condiciones; en total, $2n - 2$ condiciones.

2. *La primera y la última función deben pasar a través de los puntos extremos*. Esto agrega dos ecuaciones más:

$$a_1 x_0^2 + b_1 x_0 + c_1 = f(x_0) \tag{18.31}$$

$$a_n x_n^2 + b_n x_n + c_n = f(x_n) \tag{18.32}$$

en total tenemos $2n - 2 + 2 = 2n$ condiciones.

3. *Las primeras derivadas en los nodos interiores deben ser iguales.* La primera derivada de la ecuación (18.28) es

$$f'(x) = 2ax + b$$

Por lo tanto, de manera general la condición se representa como

$$2a_{i-1}x_{i-1} + b_{i-1} = 2a_i x_{i-1} + b_i \tag{18.33}$$

para $i = 2$ a n. Esto proporciona otras $n - 1$ condiciones, llegando a un total de $2n + n - 1 = 3n - 1$. Como se tienen $3n$ incógnitas, nos falta una condición más. A menos que tengamos alguna información adicional respecto de las funciones o sus derivadas, tenemos que realizar una elección arbitraria para calcular las constantes. Aunque hay varias opciones, elegimos la siguiente:

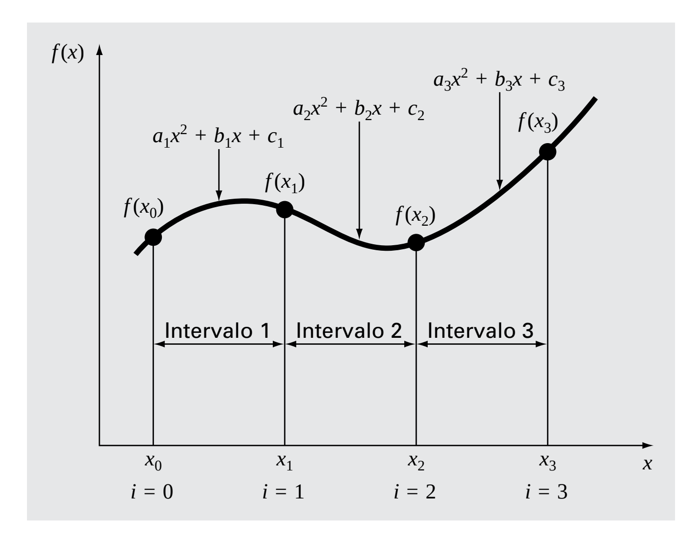


Figura 18.17
Notación utilizada para obtener trazadores cuadráticos. Observe que hay n intervalos y $n + 1$ puntos asociados con datos. El ejemplo mostrado es para $n = 3$.

4. *Suponga que en el primer punto la segunda derivada es cero*. Como la segunda derivada de la ecuación (18.28) es $2a_i$, entonces esta condición se puede expresar matemáticamente como

$$a_1 = 0 \tag{18.34}$$

La interpretación visual de esta condición es que los dos primeros puntos se unirán con una línea recta.

EJEMPLO 18.9 Trazadores cuadráticos

Planteamiento del problema Ajuste trazadores cuadráticos a los mismos datos que se utilizaron en el ejemplo 18.8 (tabla 18.1). Con los resultados estime el valor en $x = 5$.

Solución En este problema se tienen cuatro puntos asociados con datos y $n = 3$ intervalos. Por lo tanto, 3(3) = 9 incógnitas que deben determinarse. Las ecuaciones (18.29) y (18.30) dan 2(3) – 2 = 4 condiciones:

$$\begin{aligned} 20.25a_1 + 4.5b_1 + c_1 &= 1.0 \\ 20.25a_2 + 4.5b_2 + c_2 &= 1.0 \\ 49a_2 + 7b_2 + c_2 &= 2.5 \\ 49a_3 + 7b_3 + c_3 &= 2.5 \end{aligned}$$

Evaluando a la primera y la última función con los valores inicial y final, se agregan 2 ecuaciones más [ecuación (18.31)]:

$$9a_1 + 3b_1 + c_1 = 2.5$$

y [ecuación (18.32)]

$$81a_3 + 9b_3 + c_3 = 0.5$$

La continuidad de las derivadas crea adicionalmente de 3 – 1 = 2 condiciones [ecuación (18.33)]:

$$\begin{aligned} 9a_1 + b_1 &= 9a_2 + b_2 \\ 14a_2 + b_2 &= 14a_3 + b_3 \end{aligned}$$

Por último, la ecuación (18.34) determina que $a_1 = 0$. Como esta ecuación especifica a_1 de manera exacta, el problema se reduce a la solución de ocho ecuaciones simultáneas. Estas condiciones se expresan en forma matricial como

$$\begin{bmatrix} 4.5 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 20.25 & 4.5 & 1 & 0 & 0 & 0 \\ 0 & 0 & 49 & 7 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 49 & 7 & 1 \\ 3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 81 & 9 & 1 \\ 1 & 0 & -9 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 14 & 1 & 0 & -14 & -1 & 0 \end{bmatrix} \begin{Bmatrix} b_1 \\ c_1 \\ a_2 \\ b_2 \\ c_2 \\ a_3 \\ b_3 \\ c_3 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 1 \\ 2.5 \\ 2.5 \\ 2.5 \\ 0.5 \\ 0 \\ 0 \end{Bmatrix}$$

Estas ecuaciones se pueden resolver utilizando las técnicas de la parte tres, con los resultados:

$$\begin{aligned} &a_1 = 0 && b_1 = -1 && c_1 = 5.5 \\ &a_2 = 0.64 && b_2 = -6.76 && c_2 = 18.46 \\ &a_3 = -1.6 && b_3 = 24.6 && c_3 = -91.3 \end{aligned}$$

que se sustituyen en las ecuaciones cuadráticas originales para obtener la siguiente relación para cada intervalo:

$f_1(x) = -x + 5.5 \qquad 3.0 \le x \le 4.5$
$f_2(x) = 0.64x^2 - 6.76x + 18.46 \qquad 4.5 \le x \le 7.0$
$f_3(x) = -1.6x^2 + 24.6x - 91.3 \qquad 7.0 \le x \le 9.0$

Cuando se usa f_2, la predicción para $x = 5$ es, en consecuencia,

$$f_2(5) = 0.64(5)^2 - 6.76(5) + 18.46 = 0.66$$

El ajuste total por trazadores se ilustra en la figura 18.16*b*. Observe que hay dos desventajas que se alejan del ajuste: 1) la línea recta que une los dos primeros puntos y 2) el trazador para el último intervalo parece oscilar demasiado. Los trazadores cúbicos de la siguiente sección no presentan estas desventajas y, en consecuencia, son mejores métodos para la interpolación mediante trazadores.

18.6.3 Trazadores cúbicos

El objetivo en los trazadores cúbicos es obtener un polinomio de tercer grado para cada intervalo entre los nodos:

$$f_i(x) = a_i x^3 + b_i x^2 + c_i x + d_i \tag{18.35}$$

Así, para $n + 1$ puntos asociados con datos ($i = 0, 1, 2,..., n$) existen n intervalos y, en consecuencia, $4n$ incógnitas a evaluar. Como con los trazadores cuadráticos, se requieren $4n$ condiciones para evaluar las incógnitas. Éstas son:

1. Los valores de la función deben ser iguales en los nodos interiores ($2n - 2$ condiciones).
2. La primera y última función deben pasar a través de los puntos extremos (2 condiciones).
3. Las primeras derivadas en los nodos interiores deben ser iguales ($n - 1$ condiciones).
4. Las segundas derivadas en los nodos interiores deben ser iguales ($n - 1$ condiciones).
5. Las segundas derivadas en los nodos extremos son cero (2 condiciones).

La interpretación visual de la condición 5 es que la función se vuelve una línea recta en los nodos extremos. La especificación de una condición tal en los extremos nos lleva a lo que se denomina trazador “natural”. Se le da tal nombre debido a que los trazadores para el dibujo naturalmente se comportan en esta forma (figura 18.15). Si el valor de la segunda derivada en los nodos extremos no es cero (es decir, existe alguna curvatura), es posible utilizar esta información de manera alternativa para tener las dos condiciones finales.

Los cinco tipos de condiciones anteriores proporcionan el total de las $4n$ ecuaciones requeridas para encontrar los $4n$ coeficientes. Mientras es posible desarrollar trazadores cúbicos de esta forma, presentaremos una técnica alternativa que requiere la solución de sólo $n - 1$ ecuaciones. Aunque la obtención de este método (cuadro 18.3) es un poco menos directo que el de los trazadores cuadráticos, la ganancia en eficiencia bien vale la pena.

La deducción del cuadro 18.3 da como resultado la siguiente ecuación cúbica en cada intervalo:

$$\begin{aligned} f_i(x) = {} & \frac{f_i''(x_{i-1})}{6(x_i - x_{i-1})}(x_i - x)^3 + \frac{f_i''(x_i)}{6(x_i - x_{i-1})}(x - x_{i-1})^3 \\ & + \left[\frac{f(x_{i-1})}{x_i - x_{i-1}} - \frac{f''(x_{i-1})(x_i - x_{i-1})}{6}\right](x_i - x) \\ & + \left[\frac{f(x_i)}{x_i - x_{i-1}} - \frac{f''(x_i)(x_i - x_{i-1})}{6}\right](x - x_{i-1}) \end{aligned} \tag{18.36}$$

Esta ecuación contiene sólo dos incógnitas (las segundas derivadas en los extremos de cada intervalo). Las incógnitas se evalúan empleando la siguiente ecuación:

Cuadro 18.3 Obtención de trazadores cúbicos

El primer paso en la obtención (Cheney y Kincaid, 2008) se considera la observación de cómo cada par de nodos está unida por una cúbica; la segunda derivada dentro de cada intervalo es una línea recta. La ecuación (18.35) se puede derivar dos veces para verificar esta observación. Con esta base, la segunda derivada se representa mediante un polinomio de interpolación de Lagrange de primer grado [ecuación (18.22)]:

$$f_i''(x) = f_i''(x_{i-1})\frac{x - x_i}{x_{i-1} - x_i} + f_i''(x_i)\frac{x - x_{i-1}}{x_i - x_{i-1}} \tag{C18.3.1}$$

donde $f_i''(x)$ es el valor de la segunda derivada en cualquier punto x dentro del i-ésimo intervalo. Así, esta ecuación es una línea recta, que une la segunda derivada en el primer nodo $f''(x_{i-1})$ con la segunda derivada en el segundo nodo $f''(x_i)$.

Después, la ecuación (C18.3.1) se integra dos veces para obtener una expresión para $f_i(x)$. Sin embargo, esta expresión contendrá dos constantes de integración desconocidas. Dichas constantes se evalúan tomando las condiciones de igualdad de las funciones [$f(x)$ debe ser igual a $f(x_{i-1})$ en x_{i-1} y $f(x)$ debe ser igual a $f(x_i)$ en x_i)]. Al realizar estas evaluaciones, se tiene la siguiente ecuación cúbica:

$$f_i(x) = \frac{f_i''(x_{i-1})}{6(x_i - x_{i-1})}(x_i - x)^3 + \frac{f_i''(x)}{6(x_i - x_{i-1})}(x - x_{i-1})^3 + \left[\frac{f(x_{i-1})}{x_i - x_{i-1}} - \frac{f''(x_{i-1})(x_i - x_{i-1})}{6}\right](x_i - x) + \left[\frac{f(x_i)}{x_i - x_{i-1}} - \frac{f''(x_i)(x_i - x_{i-1})}{6}\right](x - x_{i-1}) \tag{C18.3.2}$$

Ahora, es claro que esta relación es una expresión mucho más compleja para un trazador cúbico para el i-ésimo intervalo que, digamos, la ecuación (18.35). Sin embargo, observe que contiene sólo dos "coeficientes" desconocidos; es decir, las segundas derivadas al inicio y al final del intervalo: $f''(x_{i-1})$ y $f''(x_i)$. De esta forma, si podemos determinar la segunda derivada en cada nodo, la ecuación (C18.3.2) es un polinomio de tercer grado que se utiliza para interpolar dentro del intervalo.

Las segundas derivadas se evalúan tomando la condición de que las primeras derivadas deben ser continuas en los nodos:

$$f_i'(x_i) = f_{i+1}'(x_i) \tag{C18.3.3}$$

La ecuación (C18.3.2) se deriva para ofrecer una expresión de la primera derivada. Si se hace esto tanto para el (i – 1)-ésimo, como para i-ésimo intervalos, y los dos resultados se igualan de acuerdo con la ecuación (C18.3.3), se llega a la siguiente relación:

$$(x_i - x_{i-1})f''(x_{i-1}) + 2(x_{i+1} - x_{i-1})f''(x_i) + (x_{i+1} - x_i)f''(x_{i+1}) = \frac{6}{x_{i+1} - x_i}[f(x_{i+1}) - f(x_i)] + \frac{6}{x_i - x_{i-1}}[f(x_{i-1}) - f(x_i)] \tag{C18.3.4}$$

Si la ecuación (C18.3.4) se escribe para todos los nodos interiores, se obtienen n – 1 ecuaciones simultáneas con n + 1 segundas derivadas desconocidas. Sin embargo, como ésta es un trazador cúbico natural, las segundas derivadas en los nodos extremos son cero y el problema se reduce a n – 1 ecuaciones con n – 1 incógnitas. Además, observe que el sistema de ecuaciones será tridiagonal. Así, no sólo se redujo el número de ecuaciones, sino que las organizamos en una forma extremadamente fácil de resolver (recuerde la sección 11.1.1).

$$(x_i - x_{i-1})f''(x_{i-1}) + 2(x_{i+1} - x_{i-1})f''(x_i) + (x_{i+1} - x_i)f''(x_{i+1}) = \frac{6}{x_{i+1} - x_i}[f(x_{i+1}) - f(x_i)] + \frac{6}{x_i - x_{i-1}}[f(x_{i-1}) - f(x_i)] \tag{18.37}$$

Si se escribe esta ecuación para todos los nodos interiores, resultan n – 1 ecuaciones simultáneas con n – 1 incógnitas. (Recuerde que las segundas derivadas en los nodos extremos son cero.) La aplicación de estas ecuaciones se ilustra con el siguiente ejemplo.

EJEMPLO 18.10 Trazadores cúbicos

Planteamiento del problema Ajuste trazadores cúbicos a los mismos datos que se usaron en los ejemplos 18.8 y 18.9 (tabla 18.1). Utilice los resultados para estimar el valor en x = 5.

Solución El primer paso consiste en usar la ecuación (18.37) para generar el conjunto de ecuaciones simultáneas que se utilizarán para determinar las segundas derivadas en los nodos. Por ejemplo, para el primer nodo interior se emplean los siguientes datos:

$x_0 = 3$ $f(x_0) = 2.5$
$x_1 = 4.5$ $f(x_1) = 1$
$x_2 = 7$ $f(x_2) = 2.5$

Estos valores se sustituyen en la ecuación (18.37):

$$(4.5-3)f''(3)+2(7-3)f''(4.5)+(7-4.5)f''(7)$$
$$=\frac{6}{7-4.5}(2.5-1)+\frac{6}{4.5-3}(2.5-1)$$

Debido a la condición de trazador natural, $f''(3) = 0$, y la ecuación se reduce a

$$8f''(4.5) + 2.5f''(7) = 9.6$$

En una forma similar, la ecuación (18.37) se aplica al segundo punto interior con el siguiente resultado:

$$2.5f''(4.5) + 9f''(7) = -9.6$$

Estas dos ecuaciones se resuelven simultáneamente:

$$f''(4.5) = 1.67909$$
$$f''(7) = -1.53308$$

Estos valores se sustituyen después en la ecuación (18.36), junto con los valores de las x y las $f(x)$, para dar

$$f_1(x)=\frac{1.67909}{6(4.5-3)}(x-3)^3+\frac{2.5}{4.5-3}(4.5-x)$$
$$+\left[\frac{1}{4.5-3}-\frac{1.67909(4.5-3)}{6}\right](x-3)$$

o

$$f_1(x) = 0.186566(x-3)^3 + 1.666667(4.5-x) + 0.246894(x-3)$$

Esta ecuación es el trazador cúbico para el primer intervalo. Se realizan sustituciones similares para tener las ecuaciones para el segundo y tercer intervalos:

$$f_2(x) = 0.111939(7-x)^3 - 0.102205(x-4.5)^3 - 0.299621(7-x)$$
$$+ 1.638783(x-4.5)$$

y

$$f_3(x) = -0.127757(9-x)^3 + 1.761027(9-x) + 0.25(x-7)$$

Las tres ecuaciones se pueden utilizar para calcular los valores dentro de cada intervalo. Por ejemplo, el valor en $x = 5$, que está dentro del segundo intervalo, se calcula como sigue

$$f_2(5) = 0.111939(7-5)^3 - 0.102205(5-4.5)^3 - 0.299621(7-5)$$
$$+ 1.638783(5-4.5) = 1.102886$$

Se calculan otros valores y los resultados se grafican en la figura 18.16*c*.

Los resultados de los ejemplos 18.8 a 18.10 se resumen en la figura 18.16. Observe cómo mejora progresivamente el ajuste conforme pasamos de trazadores lineales, a cuadráticos y cúbicos. También hemos sobrepuesto un polinomio de interpolación cúbica en la figura 18.16*c*. Aunque el trazador cúbico consiste de una serie de curvas de tercer grado, el ajuste resultante difiere del obtenido al usar un polinomio de tercer grado. Esto se debe al hecho de que el trazador natural requiere segundas derivadas iguales a cero en los nodos extremos; mientras que el polinomio cúbico no tiene tal restricción.

18.6.4 Algoritmo computacional para trazadores cúbicos

El método para calcular trazadores cúbicos, descrito en la sección anterior, es ideal para implementarse en una computadora. Recuerde que, con algunas manipulaciones inteligentes, el método se reduce a la solución de $n - 1$ ecuaciones simultáneas. Un beneficio más de la derivación es que, como lo especifica la ecuación (18.37), el sistema de ecuaciones es tridiagonal. Como se describió en la sección 11.1, existen algoritmos para resolver tales sistemas de una manera extremadamente eficiente. La figura 18.18 muestra una estructura computacional que incorpora esas características.

Observe que la rutina de la figura 18.18 da sólo un valor interpolado, *yu*, para un valor dado de la variable dependiente, *xu*. Ésta es sólo una forma en la cual se puede implementar la interpolación mediante trazadores. Por ejemplo, usted deseará determinar los coeficientes una sola vez y, después, realizar muchas interpolaciones. Además, la rutina da tanto la primera (*dy*) como la segunda derivadas (dy^2) en *xu*. Aunque no es necesario calcular esas cantidades, son útiles en muchas aplicaciones de la interpolación mediante trazadores.

18.7 INTERPOLACIÓN MULTIDIMENSIONAL

Los métodos de interpolación para problemas unidimensionales se pueden extender a la interpolación multidimensional. En esta sección se describirá el caso más sencillo de interpolación bidimensional en coordenadas cartesianas.

```
SUBROUTINE Spline (x,y,n,xu,yu,dy,d2y)
LOCAL en, fn, gn, rn, d2xn
CALL Tridiag(x,y,n,e,f,g,r)
CALL Decomp(e,f,g,n-1)
CALL Subst(e,f,g,r,n-1,d2x)
CALL Interpol(x,y,n,d2x,xu,yu,dy,d2y)
END Spline

SUBROUTINE Tridiag (x,y,n,e,f,g,r)
f1 = 2 * (x2-x0)
g1 = (x2-x1)
r1 = 6/(x2-x1) * (y2-y1)
r1 = r1+6/(x1-x0) * (y0-y1)
DOFOR i = 2, n-2
  ei = (xi-xi-1)
  fi = 2 * (xi+1 - xi-1)
  gi = (xi+1 - xi)
  ri = 6/(xi+1 - xi) * (yi+1 - yi)
  ri = ri+6/(xi - xi-1) * (yi-1 - yi)
END DO
en-1 = (xn-1 - xn-2)
fn-1 = 2 * (xn - xn-2)
rn-1 = 6/(xn - xn-1) * (yn - yn-1)
rn-1 = rn-1 + 6/(xn-1 - xn-2) * (yn-2 - yn-1)
END Tridiag
```

```
SUBROUTINE Interpol (x,y,n,d2x,xu,yu,dy,d2y)
flag = 0
i = 1
DO
  IF xu ≥ xi-1 AND xu ≤ xi THEN
     c1 = d2xi-1/6/(xi - xi-1)
     c2 = d2xi/6/(xi - xi-1)
     c3 = yi-1/(xi - xi-1) - d2xi-1 * (xi-xi-1)/6
     c4 = yi/(xi - xi-1) - d2xi * (xi-xi-1)/6
     t1 = c1 * (xi - xu)^3
     t2 = c2 * (xu - xi-1)^3
     t3 = c3 * (xi - xu)
     t4 = c4 * (xu - xi-1)
     yu = t1 + t2 + t3 + t4
     t1 = -3 * c1 * (xi - xu)^2
     t2 = 3 * c2 * (xu - xi-1)^2
     t3 = -c3
     t4 = c4
     dy = t1 + t2 + t3 + t4
     t1 = 6 * c1 * (xi - xu)
     t2 = 6 * c2 * (xu - xi-1)
     d2y = t1 + t2
     flag = 1
  ELSE
     i = i + 1
  END IF
  IF i = n + 1 OR flag = 1 EXIT
END DO
IF flag = 0 THEN
  PRINT "outside range"
  pause
END IF
END Interpol
```

Figura 18.18
Algoritmo para la interpolación mediante trazadores cúbicos.

18.7.1 Interpolación bilineal

La *interpolación bidimensional* sirve para determinar valores intermedios para funciones de dos variables, $z = f(x_i, y_i)$. Como se ilustra en la figura 18.19, se tienen valores en cuatro puntos: $f(x_1, y_1)$, $f(x_2, y_1)$, $f(x_1, y_2)$ y $f(x_2, y_2)$. Se desea interpolar entre estos puntos para estimar el valor de un punto intermedio $f(x_i, y_i)$. Si se usa una función lineal, el resultado es un plano que conecta los puntos como en la figura 18.19. Tales funciones se conocen como *bilineales*.

En la figura 18.20 se ilustra una forma sencilla de desarrollar la función bilineal. Primero, se puede mantener fijo el valor de y y aplicar interpolación lineal unidimensional en la dirección x. Usando la forma de Lagrange, el resultado en (x_i, y_1) es

$$f(x_i, y_1) = \frac{x_i - x_2}{x_1 - x_2} f(x_1, y_1) + \frac{x_i - x_1}{x_2 - x_1} f(x_2, y_1) \tag{18.38}$$

y en (x_i, y_2) es

$$f(x_i, y_2) = \frac{x_i - x_2}{x_1 - x_2} f(x_1, y_2) + \frac{x_i - x_1}{x_2 - x_1} f(x_2, y_2) \tag{18.39}$$

Estos puntos se pueden entonces usar para interpolar linealmente a lo largo de la dimensión y para obtener el resultado final,

$$f(x_i, y_i) = \frac{y_i - y_2}{y_1 - y_2} f(x_i, y_1) + \frac{y_i - y_1}{y_2 - y_1} f(x_i, y_2) \tag{18.40}$$

Figura 18.19
Ilustración gráfica de interpolación bidimensional bilineal, en la que un valor intermedio (círculo relleno) se estima con base en cuatro valores dados (círculos en blanco).

Se puede desarrollar una sola ecuación sustituyendo las ecuaciones (18.38) y (18.39) en la ecuación (18.40) para dar

$$\begin{aligned} f(x_i, y_i) = {} & \frac{x_i - x_2}{x_1 - x_2}\frac{y_i - y_2}{y_1 - y_2} f(x_1, y_1) + \frac{x_i - x_1}{x_2 - x_1}\frac{y_i - y_2}{y_1 - y_2} f(x_2, y_1) \\ & + \frac{x_i - x_2}{x_1 - x_2}\frac{y_i - y_1}{y_2 - y_1} f(x_1, y_2) + \frac{x_i - x_1}{x_2 - x_1}\frac{y_i - y_1}{y_2 - y_1} f(x_2, y_2) \end{aligned} \tag{18.41}$$

EJEMPLO 18.11 Interpolación bilineal

Planteamiento del problema Suponga que ha medido temperaturas en varias coordenadas en la superficie de una placa rectangular calentada:

$T(2, 1) = 60 \qquad T(9, 1) = 57.5$

$T(2, 6) = 55 \qquad T(9, 6) = 70$

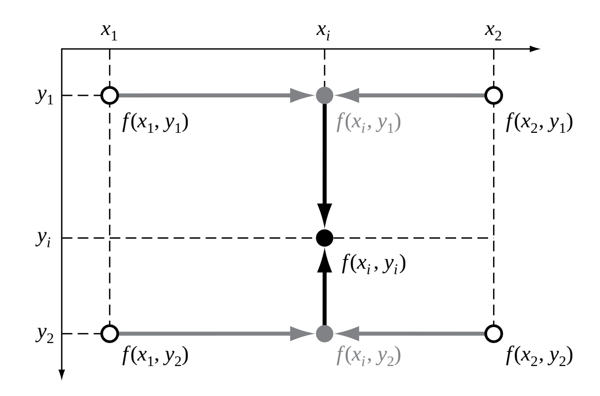


Figura 18.20
La interpolación bidimensional bilineal se puede implementar aplicando primero interpolación lineal unidimensional a lo largo de la dimensión x para determinar valores de x_i. Estos valores se pueden entonces usar para interpolar linealmente a lo largo de la dimensión y para obtener el resultado final en x_i, y_i.

Use interpolación bilineal para estimar la temperatura en $x_i = 5.25$ y $y_i = 4.8$.

Solución Al sustituir estos valores en la ecuación (18.41) se obtiene

$$f(5.5, 4) = \frac{5.25-9}{2-9}\frac{4.8-6}{1-6}60 + \frac{5.25-2}{9-2}\frac{4.8-6}{1-6}57.5 + \frac{5.25-9}{2-9}\frac{4.8-1}{6-1}55 + \frac{5.25-2}{9-2}\frac{4.8-1}{6-1}70 = 61.2143$$

Observe que más allá de la sencilla interpolación bilineal descrita en el ejemplo anterior, también se pueden usar polinomios de grado más alto y trazadores para interpolar en dos dimensiones. Además, estos métodos se pueden extender fácilmente a tres dimensiones. Se volverá a este tema en la reseña de aplicaciones de software para interpolación al final del capítulo 19.

PROBLEMAS

18.1 Estime el logaritmo natural de 10 por medio de interpolación lineal.

a) Interpole entre log 8 = 0.9030900 y log 12 = 1.0791812.
b) Interpole entre log 9 = 0.9542425 y log 11 = 1.0413927. Para cada una de las interpolaciones calcule el error relativo porcentual con base en el valor verdadero.

18.2 Ajuste un polinomio de interpolación de Newton de segundo grado para estimar el log 10, con los datos del problema 18.1 en x = 8, 9 y 11. Calcule el error relativo porcentual verdadero.

18.3 Ajuste un polinomio de interpolación de Newton de tercer grado para estimar log 10 con los datos del problema 18.1.

18.4 Repita los problemas 18.1 a 18.3 usando el polinomio de Lagrange.

18.5 Dados los datos

x	1.6	2	2.5	3.2	4	4.5
$f(x)$	2	8	14	15	8	2

a) Calcule $f(2.8)$ con el uso de polinomios de interpolación de Newton de grados 1 a 3. Elija la secuencia de puntos más apropiada para alcanzar la mayor exactitud posible para sus estimaciones.
b) Utilice la ecuación (18.18) para estimar el error de cada predicción.

18.6 Dados los datos

x	1	2	3	5	7	8
$f(x)$	3	6	19	99	291	444

Calcule $f(4)$ con el uso de polinomios de interpolación de Newton de grados 1 a 4. Elija los puntos base para obtener una buena exactitud. ¿Qué indican los resultados en relación con el grado del polinomio que se emplea para generar los datos de la tabla?

18.7 Vuelva a hacer el problema 18.6 con el uso de polinomios de Lagrange de grados 1 a 3.

18.8 Los siguientes datos provienen de una tabla y fueron medidos con alta precisión. Use el mejor método numérico (para este tipo de problema) para determinar y en x = 3.5. Observe que un polinomio dará un valor exacto. Su solución debe probar que su resultado es exacto.

x	0	1.8	5	6	8.2	9.2	12
y	26	16.415	5.375	3.5	2.015	2.54	8

18.9 Use el polinomio de interpolación de Newton para determinar y en x = 3.5 con la mayor exactitud posible. Calcule las diferencias divididas finitas como en la figura 18.5 y ordene sus puntos para obtener exactitud óptima y convergencia.

x	0	1	2.5	3	4.5	5	6
y	2	5.4375	7.3516	7.5625	8.4453	9.1875	12

18.10 Use el polinomio de interpolación de Newton para determinar y en x = 8 con la mayor exactitud posible. Calcule las diferencias divididas finitas como en la figura 18.5 y ordene sus puntos para obtener exactitud óptima y convergencia.

x	0	1	2	5.5	11	13	16	18
y	0.5	3.134	5.3	9.9	10.2	9.35	7.2	6.2

18.11 Emplee interpolación inversa con el uso de un polinomio de interpolación cúbico y de bisección, para determinar el valor de x que corresponde a $f(x)$ = 0.23, para los datos tabulados que siguen:

x	2	3	4	5	6	7
y	0.5	0.3333	0.25	0.2	0.1667	1.1429

18.12 Utilice interpolación inversa para determinar el valor de x que corresponde a $f(x)$ = 0.85, para los datos tabulados siguientes:

x	0	1	2	3	4	5
$f(x)$	0	0.5	0.8	0.9	0.941176	0.961538

Observe que los valores de la tabla se generaron con la función $f(x) = x^2/(1 + x^2)$.

a) Determine en forma analítica el valor correcto.
b) Use interpolación cúbica de x *versus* y.
c) Utilice interpolación inversa con interpolación cuadrática y la fórmula cuadrática.

d) Emplee interpolación inversa con interpolación cúbica y bisección. Para los incisos *b*) a *d*) calcule el error relativo porcentual verdadero.

18.13 Desarrolle trazadores cuadráticos para los cinco primeros puntos asociados con datos del problema 18.5, y pronostique $f(3.4)$ y $f(2.2)$.

18.14 Obtenga trazadores cúbicos para los datos del problema 18.6, y *a*) pronostique $f(4)$ y $f(2.5)$, y *b*) verifique que $f_2(3)$ y $f_3(3) = 19$.

18.15 Determine los coeficientes de la parábola que pasa por los últimos tres puntos del problema 18.5.

18.16 Determine los coeficientes de la ecuación cúbica que pasa por los primeros cuatro puntos del problema 18.6.

18.17 Desarrolle, depure y pruebe un programa en cualquier lenguaje de alto nivel o de macros que elija, para implantar la interpolación de polinomios de Newton, con base en la figura 18.7.

18.18 Pruebe el programa que desarrolló en el problema 18.17 con la duplicación del cálculo del ejemplo 18.5.

18.19 Use el programa que desarrolló en el problema 18.17 para resolver los problemas 18.1 a 18.3.

18.20 Utilice el programa que desarrolló en el problema 18.17 para solucionar los problemas 18.5 y 18.6. Utilice todos los datos para desarrollar polinomios de primero a quinto grados. Para ambos problemas, haga la gráfica del error estimado *versus* el grado.

18.21 Desarrolle, depure y pruebe un programa en el lenguaje de alto nivel o macros que elija, para implantar la interpolación de Lagrange. Utilice como base el pseudocódigo de la figura 18.11. Pruébelo con la duplicación del ejemplo 18.7.

18.22 Una aplicación útil de la interpolación de Lagrange se denomina *búsqueda en la tabla*. Como el nombre lo indica, involucra "buscar" un valor intermedio en una tabla. Para desarrollar dicho algoritmo, en primer lugar se almacena la tabla de los valores de x y $f(x)$ en un par de arreglos unidimensionales. Después, dichos valores se pasan a una función junto con el valor de x que se desea evaluar. La función hace luego dos tareas. En primer lugar, hace un ciclo hacia abajo de la tabla hasta que encuentra el intervalo en el que se localiza la incógnita. Después aplica una técnica como la interpolación de Lagrange para determinar el valor apropiado de $f(x)$. Desarrolle una función así con el uso de un polinomio cúbico de Lagrange para ejecutar la interpolación. Para intervalos intermedios ésta es una buena elección porque la incógnita se localiza en el intervalo a la mitad de los cuatro puntos necesarios para generar la expresión cúbica. Para los intervalos primero y último, use un polinomio cuadrático de Lagrange. Asimismo, haga que el código detecte cuando el usuario pida un valor fuera del rango de las x. Para esos casos, la función debe desplegar un mensaje de error. Pruebe su programa para $f(x) = \ln x$ con los datos $x = 1, 2, \ldots, 10$.

18.23 Desarrolle, depure y pruebe un programa en cualquier lenguaje de alto nivel o de macros de su elección, para implantar la interpolación con splin cúbico con base en la figura 18.18. Pruebe el programa con la repetición del ejemplo 18.10.

18.24 Emplee el software desarrollado en el problema 18.23 para ajustar trazadores cúbicos para los datos de los problemas 18.5 y 18.6. Para ambos casos, pronostique $f(2.25)$.

18.25 Emplee la porción de la tabla de vapor que se da para el H_2O sobrecalentado a 200 MPa, para *a*) encontrar la entropía correspondiente s para un volumen específico v de 0.108 m^3/kg con interpolación lineal, *b*) encontrar la misma entropía correspondiente con el uso de interpolación cuadrática, y *c*) hallar el volumen correspondiente a una entropía de 6.6 con el empleo de interpolación inversa.

v (m^3/kg)	0.10377	0.11144	0.1254
s (kJ/kg · K)	6.4147	6.5453	6.7664

18.26 La función de Runge se escribe como

$$f(x) = \frac{1}{1 + 25x^2}$$

a) Desarrolle una gráfica de esta función para el intervalo de $x = -1$ a 1.
b) Genere y grafique el polinomio de interpolación de Lagrange usando valores de función equiespaciados correspondientes a $x = -1, -0.5, 0, 0.5$ y 1.
c) Use los cinco puntos de *b*) para estimar $f(0.8)$ con polinomios de interpolación de Newton de grado primero a cuarto.
d) Genere y grafique un trazador cúbico usando los cinco puntos de *b*).
e) Explique sus resultados.

18.27 La siguiente es la función incorporada de gibas que usa MATLAB para demostrar algunas de sus capacidades numéricas:

$$f(x) = \frac{1}{(x - 0.3)^2 + 0.01} + \frac{1}{(x - 0.9)^2 + 0.04} - 6$$

La función de gibas muestra regiones tanto llanas como escarpadas a lo largo de un rango relativamente corto de x. Genere valores de esta función a intervalos de 0.1 en el rango de $x = 0$ a 1. Ajuste estos datos con un trazador cúbico y cree una gráfica que compare el ajuste con la función exacta de gibas.

18.28 Los siguientes datos definen la concentración de oxígeno disuelto a nivel del mar para agua dulce como función de la temperatura:

T, °C	0	8	16	24	32	40
o, mg/L	14.621	11.843	9.870	8.418	7.305	6.413

Estime $o(27)$ usando: *a*) interpolación lineal, *b*) polinomio de interpolación de Newton y *c*) trazadores cúbicos. Observe que el resultado exacto es 7.986 mg/L.

18.29 Genere ocho puntos igualmente espaciados de la función

$$f(t) = \operatorname{sen}^2 t$$

De $t = 0$ a 2π. Ajuste estos datos con *a*) un polinomio de interpolación de séptimo grado y *b*) un trazador cúbico.

18.30 Se miden temperaturas en varios puntos de una placa calentada (tabla P18.30). Estime la temperatura en *a*) $x = 4$, $y = 3.2$ y *b*) $x = 4.3$, $y = 2.7$.

Tabla P18.30 Temperatura (°C) en varios puntos de una placa calentada cuadrada.

	$x = 0$	$x = 2$	$x = 4$	$x = 6$	$x = 8$
$y = 0$	100.00	90.00	80.00	70.00	60.00
$y = 2$	85.00	64.49	53.50	48.15	50.00
$y = 4$	70.00	48.90	38.43	35.03	40.00
$y = 6$	55.00	38.78	30.39	27.07	30.00
$y = 8$	40.00	35.00	30.00	25.00	20.00

CAPÍTULO 19

Aproximación de Fourier

Hasta aquí, en nuestra presentación de la interpolación se han destacado los polinomios estándar, es decir, las combinaciones lineales de los monomios 1, x, x^2,…, x^m (figura 19.1*a*). Ahora veremos otra clase de funciones que son trascendentales en la ingeniería. Éstas son las funciones trigonométricas 1, cos x, cos $2x$,…, cos nx, sen x, sen $2x$,…, sen nx (figura 19.1*b*).

Los ingenieros a menudo tratan con sistemas que oscilan o vibran. Como es de esperarse, las funciones trigonométricas juegan un papel importante en el modelado de tales problemas. La *aproximación de Fourier* representa un esquema sistemático para utilizar series trigonométricas con este propósito.

Una de las características distintivas del análisis de Fourier es que trata con los dominios de la frecuencia y del tiempo. Como algunos ingenieros requieren trabajar con el último, se ha dedicado gran parte del siguiente material a ofrecer una visión general de la aproximación de Fourier. Un aspecto clave de esta visión será familiarizarse con el dominio de la frecuencia. Luego de dicha orientación se presenta una introducción a los métodos numéricos para calcular transformadas de Fourier discretas.

19.1 AJUSTE DE CURVAS CON FUNCIONES SINUSOIDALES

Una función periódica $f(t)$ es aquella para la cual

$$f(t) = f(t + T) \tag{19.1}$$

donde T es una constante llamada el *periodo*, que es el valor menor para el cual es válida la ecuación (19.1). Entre los ejemplos comunes se encuentran diversas formas de onda, como ondas cuadradas y dientes de sierra (figura 19.2). Las ondas fundamentales son las funciones sinusoidales.

En el presente análisis se usará el término *sinusoide* para representar cualquier forma de onda que se pueda describir como un seno o un coseno. No existe una convención muy clara para elegir entre estas funciones y, en cualquier caso, los resultados serán idénticos. En este capítulo se usará el coseno, que generalmente se expresa como

$$f(t) = A_0 + C_1 \cos(\omega_0 t + \theta) \tag{19.2}$$

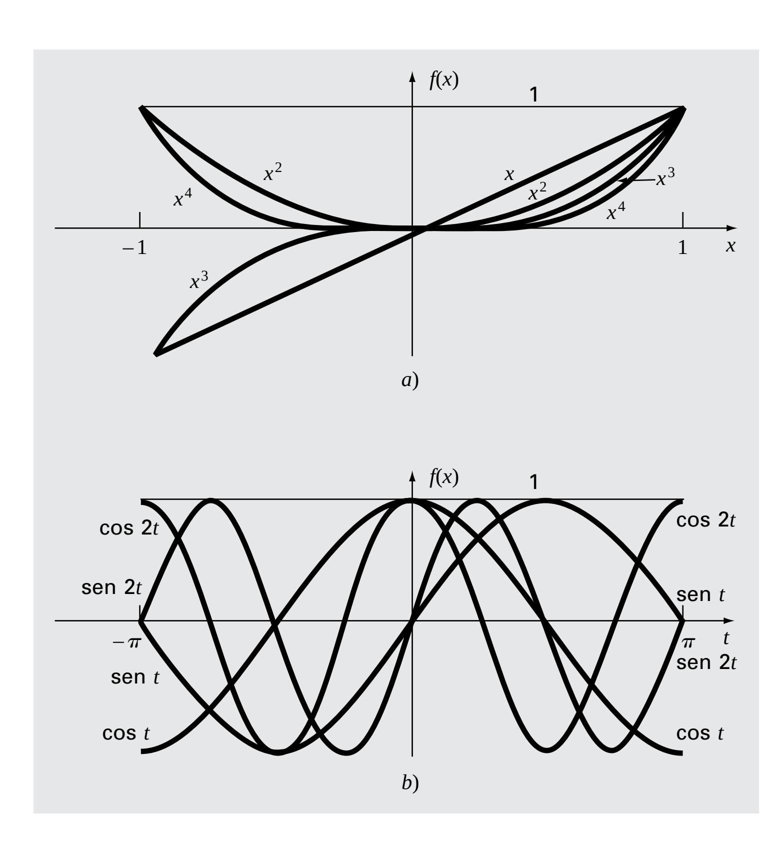


Figura 19.1
a) Los primeros cinco monomios y *b*) funciones trigonométricas. Observe que en los intervalos mostrados, ambos tipos de funciones están en el rango de –1 a 1. Sin embargo, advierta que los valores pico de los monomios se presentan todos en los extremos; mientras que en las funciones trigonométricas los picos están uniformemente distribuidos en todo el intervalo.

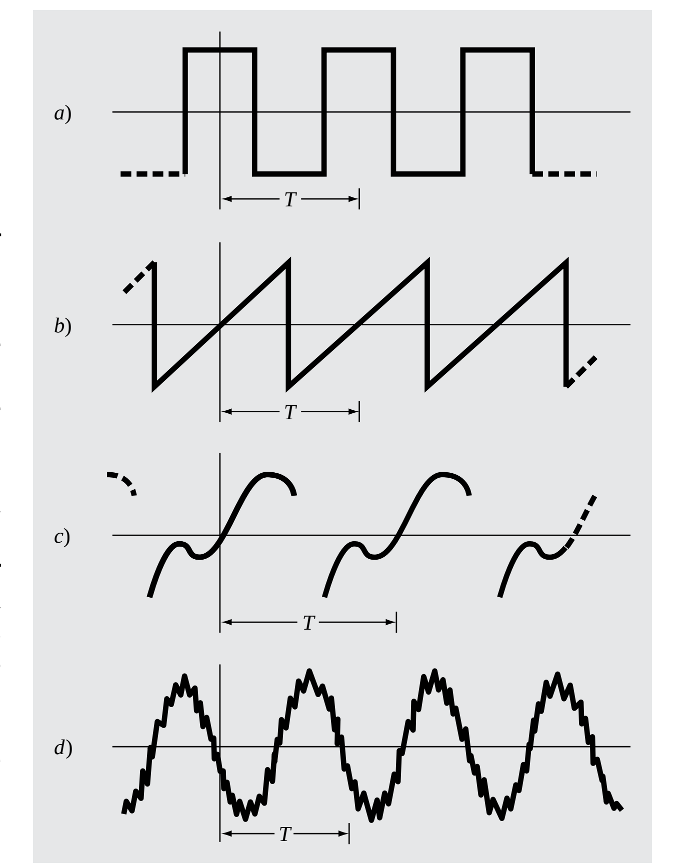


Figura 19.2
Además de las funciones trigonométricas seno y coseno, las funciones periódicas comprenden formas de onda como *a*) la onda cuadrada y *b*) la onda dientes de sierra. Más allá de estas formas idealizadas, las señales periódicas en la naturaleza pueden ser *c*) no ideales y *d*) contaminadas por ruido. Las funciones trigonométricas sirven para representar y analizar todos estos casos.

Así, cuatro parámetros sirven para caracterizar la sinusoide (figura 19.3). El *valor medio* A_0, establece la altura promedio sobre las abscisas. La *amplitud* C_1 especifica la altura de la oscilación. La *frecuencia angular* ω_0 caracteriza con qué frecuencia se presentan los ciclos. Finalmente, el ángulo de fase, o *corrimiento de fase* θ, parametriza en qué extensión la sinusoide está corrida horizontalmente. Esto puede medirse como la distancia en radianes desde $t = 0$ hasta el punto donde la función coseno empieza un nuevo ciclo. Como se ilustra en la figura 19.4*a*, un valor negativo se conoce como un *ángulo de fase de atraso*, ya que la curva $\cos(\omega_0 t - \theta)$ comienza un nuevo ciclo de θ radianes después del $\cos(\omega_0 t)$. Así, se dice que $\cos(\omega_0 t - \theta)$ tiene un retraso $\cos(\omega_0 t)$. En forma opuesta, como se muestra en la figura 19.4*b*, un valor positivo se refiere como un ángulo de fase de adelanto.

Observe que la frecuencia angular (en radianes/tiempo) se relaciona con la frecuencia f (en ciclos/tiempo) mediante

$$\omega_0 = 2\pi f \tag{19.3}$$

y, a su vez, la frecuencia está relacionada con el periodo T (en unidades de tiempo) mediante

$$f = \frac{1}{T} \tag{19.4}$$

Aunque la ecuación (19.2) representa una caracterización matemática adecuada de una sinusoide, es difícil trabajar desde el punto de vista del ajuste de curvas, pues el corrimiento de fase está incluido en el argumento de la función coseno. Esta deficiencia se resuelve empleando la identidad trigonométrica

$$C_1 \cos(\omega_0 t + \theta) = C_1[\cos(\omega_0 t)\cos(\theta) - \operatorname{sen}(\omega_0 t)\operatorname{sen}(\theta)] \tag{19.5}$$

Sustituyendo la ecuación (19.5) en la (19.2) y agrupando términos se obtiene (figura 19.3*b*)

$$f(t) = A_0 + A_1 \cos(\omega_0 t) + B_1 \operatorname{sen}(\omega_0 t) \tag{19.6}$$

donde

$$A_1 = C_1 \cos(\theta) \qquad B_1 = -C_1 \operatorname{sen}(\theta) \tag{19.7}$$

Dividiendo las dos ecuaciones anteriores y despejando se obtiene

$$\theta = \arctan\left(-\frac{B_1}{A_1}\right) \tag{19.8}$$

donde, si $A_1 < 0$, sume π a θ. Si se elevan al cuadrado y se suman las ecuaciones (19.7) llegaríamos a

$$C_1 = \sqrt{A_1^2 + B_1^2} \tag{19.9}$$

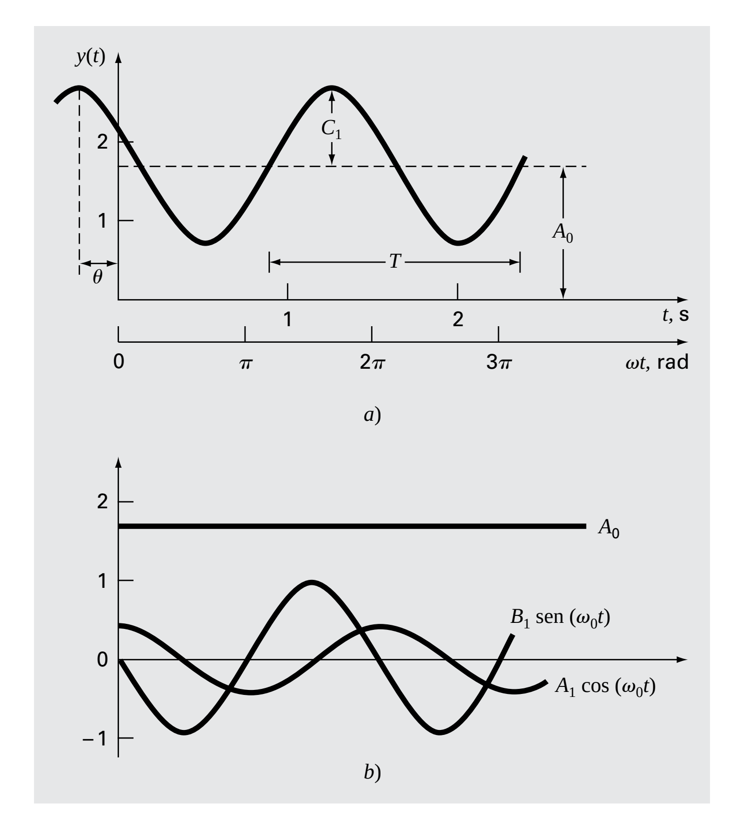


Figura 19.3
a) Una gráfica de la función sinusoidal $y(t) = A_0 + C_1 \cos(\omega_0 t + \theta)$. En este caso, $A_0 = 1.7$, $C_1 = 1$, $\omega_0 = 2\pi/T = 2\pi/(1.5 \text{ s})$, y $\theta = \pi/3$ radianes = 1.0472 (= 0.25 s). Otros parámetros que se utilizan para describir la curva son la frecuencia $f = \omega_0/(2\pi)$, que en este caso es 1 ciclo/(1.5 s), y el periodo $T = 1.5$ s. *b)* Una expresión alternativa para la misma curva es $y(t) = A_0 + A_1 \cos(\omega_0 t) + B_1 \text{sen}(\omega_0 t)$. Los tres componentes de esta función se ilustran en *b)*, donde $A_1 = 0.5$ y $B_1 = -0.866$. La suma de las tres curvas en *b)* da como resultado la curva simple en *a)*.

Así, la ecuación (19.6) representa una fórmula alternativa de la ecuación (19.2) que también requiere cuatro parámetros; pero que se encuentra en el formato de un modelo lineal general [recuerde la ecuación (17.23)]. Como se analizará en la próxima sección, es posible aplicarlo simplemente como base para un ajuste por mínimos cuadrados.

Sin embargo, antes de iniciar con la próxima sección, se deberá resaltar que se puede haber empleado la función seno en lugar de coseno, como modelo fundamental de la ecuación (19.2). Por ejemplo,

$$f(t) = A_0 + C_1 \text{sen}(\omega_0 t + \delta)$$

se pudo haber usado. Se aplican relaciones simples para convertir una forma en otra:

$$\text{sen}(\omega_0 t + \delta) = \cos\left(\omega_0 t + \delta - \frac{\pi}{2}\right)$$

y

$$\cos(\omega_0 t + \theta) = \text{sen}\left(\omega_0 t + \theta + \frac{\pi}{2}\right) \tag{19.10}$$

En otras palabras, $\theta = \delta - \pi/2$. La única consideración importante es que se debe usar una u otra forma de manera consistente. Aquí, usaremos la versión coseno en todo el análisis.

19.1.1 Ajuste por mínimos cuadrados de una sinusoide

La ecuación (19.6) se entiende como un modelo lineal por mínimos cuadrados

$$y = A_0 + A_1 \cos(\omega_0 t) + B_1 \text{sen}(\omega_0 t) + e \tag{19.11}$$

que es sólo otro ejemplo del modelo general [recuerde la ecuación (17.23)]

$$y = a_0 z_0 + a_1 z_1 + a_2 z_2 + \cdots + a_m z_m + e \tag{17.23}$$

donde $z_0 = 1$, $z_1 = \cos(\omega_0 t)$, $z_2 = \text{sen}(\omega_0 t)$ y todas las otras $z = 0$. Así, nuestro objetivo es determinar los valores de los coeficientes que minimicen la función

$$S_r = \sum_{i=1}^{N} \{y_i - [A_0 + A_1 \cos(\omega_0 t_i) + B_1 \text{sen}(\omega_0 t_i)]\}^2$$

Las ecuaciones normales para lograr esta minimización se expresan en forma matricial como [recuerde la ecuación (17.25)]

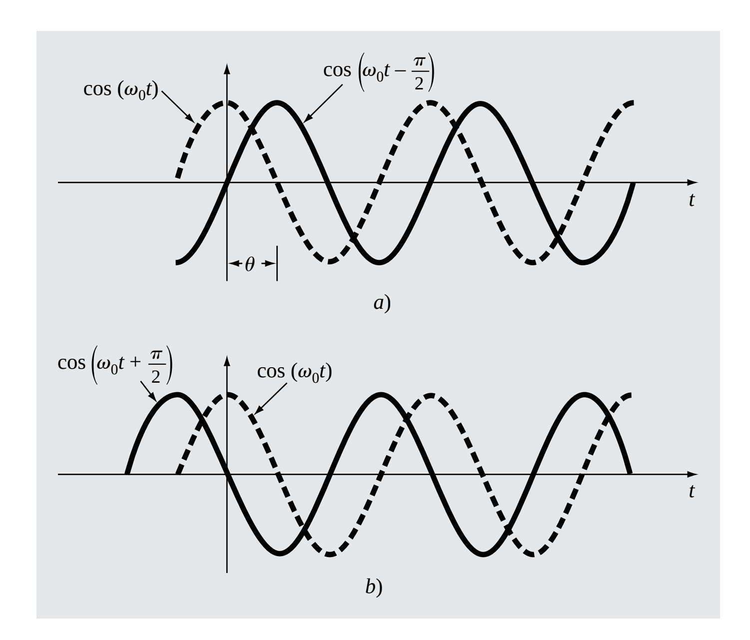


Figura 19.4
Representaciones gráficas de *a*) un ángulo de fase de atraso y *b*) un ángulo de fase de adelanto. Observe que la curva atrasada en *a*) puede describirse de manera alternativa como $\cos(\omega_0 t + 3\pi/2)$. En otras palabras, si una curva se atrasa en un ángulo α, también se puede representar como adelanto en $2\pi - \alpha$.

$$\begin{bmatrix} N & \sum\cos(\omega_0 t) & \sum\text{sen}(\omega_0 t) \\ \sum\cos(\omega_0 t) & \sum\cos^2(\omega_0 t) & \sum\cos(\omega_0 t)\ \text{sen}(\omega_0 t) \\ \sum\text{sen}(\omega_0 t) & \sum\cos(\omega_0 t)\ \text{sen}(\omega_0 t) & \sum\text{sen}^2(\omega_0 t) \end{bmatrix} \begin{Bmatrix} A_0 \\ A_1 \\ B_1 \end{Bmatrix} = \begin{Bmatrix} \sum y \\ \sum y\cos(\omega_0 t) \\ \sum y\ \text{sen}(\omega_0 t) \end{Bmatrix} \tag{19.12}$$

Estas ecuaciones sirven para encontrar los coeficientes desconocidos. Aunque, en lugar de hacer esto, se examina el caso especial donde hay N observaciones espaciadas de manera uniforme a intervalos Δt y con una longitud total $T = (N - 1)\Delta t$. En esta situación, se determinan los siguientes valores promedio (véase el problema 19.1):

$$\begin{aligned} &\frac{\sum\text{sen}(\omega_0 t)}{N} = 0 \qquad \frac{\sum\cos(\omega_0 t)}{N} = 0 \\ &\frac{\sum\text{sen}^2(\omega_0 t)}{N} = \frac{1}{2} \qquad \frac{\sum\cos^2(\omega_0 t)}{N} = \frac{1}{2} \\ &\frac{\sum\cos(\omega_0 t)\ \text{sen}(\omega_0 t)}{N} = 0 \end{aligned} \tag{19.13}$$

Así, para los puntos igualmente espaciados, las ecuaciones normales se convierten en

$$\begin{bmatrix} N & 0 & 0 \\ 0 & N/2 & 0 \\ 0 & 0 & N/2 \end{bmatrix} \begin{Bmatrix} A_0 \\ A_1 \\ B_1 \end{Bmatrix} = \begin{Bmatrix} \sum y \\ \sum y\cos(\omega_0 t) \\ \sum y\ \text{sen}(\omega_0 t) \end{Bmatrix}$$

La inversa de una matriz diagonal es simplemente otra matriz diagonal, cuyos elementos son los recíprocos de la matriz original. Así, los coeficientes se determinan como

$$\begin{Bmatrix} A_0 \\ A_1 \\ B_1 \end{Bmatrix} = \begin{bmatrix} 1/N & 0 & 0 \\ 0 & 2/N & 0 \\ 0 & 0 & 2/N \end{bmatrix} \begin{Bmatrix} \sum y \\ \sum y\cos(\omega_0 t) \\ \sum y\ \text{sen}(\omega_0 t) \end{Bmatrix}$$

o

$$A_0 = \frac{\sum y}{N} \tag{19.14}$$

$$A_1 = \frac{2}{N}\sum y\cos(\omega_0 t) \tag{19.15}$$

$$B_1 = \frac{2}{N}\sum y\,\text{sen}(\omega_0 t) \tag{19.16}$$

EJEMPLO 19.1 Ajuste por mínimos cuadrados a una sinusoide

Planteamiento del problema La curva de la figura 19.3 se describe por $y = 1.7 + \cos(4.189t + 1.0472)$. Genere 10 valores discretos para esta curva a intervalos $\Delta t = 0.15$ en el intervalo de $t = 0$ a $t = 1.35$. Utilice esta información para evaluar los coeficientes de la ecuación (19.11) mediante un ajuste por mínimos cuadrados.

Solución Los datos requeridos para evaluar los coeficientes con $\omega = 4.189$ son

t	y	$y\cos(\omega_0 t)$	$y\,\text{sen}(\omega_0 t)$
0	2.200	2.200	0.000
0.15	1.595	1.291	0.938
0.30	1.031	0.319	0.980
0.45	0.722	–0.223	0.687
0.60	0.786	–0.636	0.462
0.75	1.200	–1.200	0.000
0.90	1.805	–1.460	–1.061
1.05	2.369	–0.732	–2.253
1.20	2.678	0.829	–2.547
1.35	2.614	2.114	–1.536
Σ =	17.000	2.502	–4.330

Estos resultados se utilizan para determinar [ecuaciones (19.14) a (19.16)]

$$A_0 = \frac{17.000}{10} = 1.7 \qquad A_1 = \frac{2}{10}2.502 = 0.500 \qquad B_1 = \frac{2}{10}(-4.330) = -0.866$$

De esta manera, el ajuste por mínimos cuadrados es

$$y = 1.7 + 0.500\cos(\omega_0 t) - 0.866\,\text{sen}(\omega_0 t)$$

El modelo se expresa también en el formato de la ecuación (19.2) calculando [ecuación (19.8)]

$$\theta = \arctan\left(-\frac{-0.866}{0.500}\right) = 1.0472$$

y [ecuación (19.9)]

$$C_1 = \sqrt{(0.5)^2 + (-0.866)^2} = 1.00$$

cuyo resultado es

$$y = 1.7 + \cos(\omega_0 t + 1.0472)$$

o, en forma alternativa, con seno utilizando la ecuación (19.10)

$$y = 1.7 + \text{sen}(\omega_0 t + 2.618)$$

El análisis anterior se puede extender al modelo general

$$f(t) = A_0 + A_1 \cos(\omega_0 t) + B_1 \operatorname{sen}(\omega_0 t) + A_2 \cos(2\omega_0 t) + B_2 \operatorname{sen}(2\omega_0 t) + \cdots + A_m \cos(m\omega_0 t) + B_m \operatorname{sen}(m\omega_0 t)$$

donde, para datos igualmente espaciados, los coeficientes se evalúan con

$$A_0 = \frac{\sum y}{N}$$

$$\left.\begin{aligned} A_j &= \frac{2}{N}\sum y \cos(j\omega_0 t) \\ B_j &= \frac{2}{N}\sum y \operatorname{sen}(j\omega_0 t) \end{aligned}\right\} \quad j = 1, 2, \ldots, m$$

Aunque estas relaciones se utilizan para ajustar datos en el sentido de la regresión (es decir, $N > 2m + 1$), una aplicación alternativa es emplearlos para la interpolación o colocación (es decir, usarlos en el caso donde el número de incógnitas, $2m + 1$, es igual al número de datos, N). Éste es el procedimiento usado en la serie de Fourier continua, como se estudiará a continuación.

19.2 SERIE DE FOURIER CONTINUA

En el curso del estudio de problemas de flujo de calor, con el análisis de Fourier se demostró que una función periódica arbitraria se representa por medio de una serie infinita de sinusoides con frecuencias relacionadas de manera armónica. Para una función con un periodo T, se escribe una serie de Fourier continua[1]

$$f(t) = a_0 + a_1 \cos(\omega_0 t) + b_1 \operatorname{sen}(\omega_0 t) + a_2 \cos(2\omega_0 t) + b_2 \operatorname{sen}(2\omega_0 t) + \cdots$$

o, de manera concisa,

$$f(t) = a_0 + \sum_{k=1}^{\infty} [a_k \cos(k\omega_0 t) + b_k \operatorname{sen}(k\omega_0 t)] \tag{19.17}$$

donde $\omega_0 = 2\pi/T$ se denomina la *frecuencia fundamental* y sus múltiplos constantes $2\omega_0$, $3\omega_0$, etc., se denominan *armónicos*. De esta forma, la ecuación (19.17) expresa a $f(t)$ como una combinación lineal de las funciones base: 1, $\cos(\omega_0 t)$, $\operatorname{sen}(\omega_0 t)$, $\cos(2\omega_0 t)$, $\operatorname{sen}(2\omega_0 t)$,…

Como se describe en el cuadro 19.1, los coeficientes de la ecuación (19.17) se calculan por medio de

$$a_k = \frac{2}{T}\int_0^T f(t)\cos(k\omega_0 t)\,dt \tag{19.18}$$

y

$$b_k = \frac{2}{T}\int_0^T f(t)\operatorname{sen}(k\omega_0 t)\,dt \tag{19.19}$$

para $k = 1, 2,\ldots$ y

$$a_0 = \frac{1}{T}\int_0^T f(t)\,dt \tag{19.20}$$

[1] La existencia de las series de Fourier está referida en las condiciones de Dirichlet, las cuales especifican que la función periódica tiene un número finito de máximos y mínimos, y que hay un número finito de saltos discontinuos. En general, todas las funciones periódicas obtenidas físicamente satisfacen tales condiciones.

Cuadro 19.1 Determinación de los coeficientes de la serie de Fourier continua

Como se hizo para los datos discretos de la sección 19.1.1, se establecen las siguientes relaciones:

$$\int_0^T \text{sen}(k\omega_0 t)\, dt = \int_0^T \cos(k\omega_0 t)\, dt = 0 \qquad \text{(C19.1.1)}$$

$$\int_0^T \cos(k\omega_0 t)\, \text{sen}(g\omega_0 t)\, dt = 0 \qquad \text{(C19.1.2)}$$

$$\int_0^T \text{sen}(k\omega_0 t)\, \text{sen}(g\omega_0 t)\, dt = 0 \qquad \text{(C19.1.3)}$$

$$\int_0^T \cos(k\omega_0 t)\, \cos(g\omega_0 t)\, dt = 0 \qquad \text{(C19.1.4)}$$

$$\int_0^T \text{sen}^2(k\omega_0 t)\, dt = \int_0^T \cos^2(k\omega_0 t)\, dt = \frac{T}{2} \qquad \text{(C19.1.5)}$$

Para evaluar los coeficientes, cada lado de la ecuación (19.17) se integra obteniéndose

$$\int_0^T f(t)\, dt = \int_0^T a_0\, dt + \int_0^T \sum_{k=1}^{\infty} [a_k \cos(k\omega_0 t) + b_k\, \text{sen}(k\omega_0 t)]\, dt$$

Como cada término en la sumatoria es de la forma de la ecuación (C19.1.1), la ecuación se convierte en

$$\int_0^T f(t)\, dt = a_0 T$$

en la cual se despeja para tener

$$a_0 = \frac{\int_0^T f(t)\, dt}{T}$$

Así, a_0 es simplemente el valor medio de la función a lo largo del periodo.

Para evaluar uno de los coeficientes del coseno, por ejemplo, a_m, la ecuación (19.17) se multiplica por $(m\omega_0 t)$ e integra para dar

$$\int_0^T f(t)\cos(m\omega_0 t)\, dt = \int_0^T a_0 \cos(m\omega_0 t)\, dt + \int_0^T \sum_{k=1}^{\infty} a_k \cos(k\omega_0 t)\, \cos(m\omega_0 t)\, dt + \int_0^T \sum_{k=1}^{\infty} b_k\, \text{sen}(k\omega_0 t)\, \cos(m\omega_0 t)\, dt \qquad \text{(C19.1.6)}$$

En las ecuaciones (C19.1.1), (C19.1.2) y (C19.1.4) se observa que todos los términos del lado derecho son cero, con excepción del caso donde $k = m$. Este último caso se puede evaluar con la ecuación (C19.1.5) y, por lo tanto, de la ecuación (C19.1.6) se obtiene a_m, o de manera más general [ecuación (19.18)],

$$a_k = \frac{2}{T}\int_0^T f(t)\cos(k\omega_0 t)\, dt$$

para $k = 1, 2, \ldots$

En forma similar, la ecuación (19.17) se multiplica por $\text{sen}(m\omega_0 t)$, se integra y se manipula para dar la ecuación (19.19).

EJEMPLO 19.2 Aproximación de la serie de Fourier continua

Planteamiento del problema Utilice la serie de Fourier continua para aproximar la función de onda cuadrada o rectangular (figura 19.5)

$$f(t) = \begin{cases} -1 & -T/2 < t < -T/4 \\ 1 & -T/4 < t < T/4 \\ -1 & T/4 < t < T/2 \end{cases}$$

Solución Como la altura promedio de la onda es cero, se obtiene en forma directa un valor de $a_0 = 0$. Los coeficientes restantes se evalúan como sigue [ecuación (19.18)]

$$a_k = \frac{2}{T}\int_{-T/2}^{T/2} f(t)\cos(k\omega_0 t)\, dt$$

$$= \frac{2}{T}\left[-\int_{-T/2}^{-T/4}\cos(k\omega_0 t)\, dt + \int_{-T/4}^{T/4}\cos(k\omega_0 t)\, dt - \int_{T/4}^{T/2}\cos(k\omega_0 t)\, dt\right]$$

Las integrales se evalúan para dar

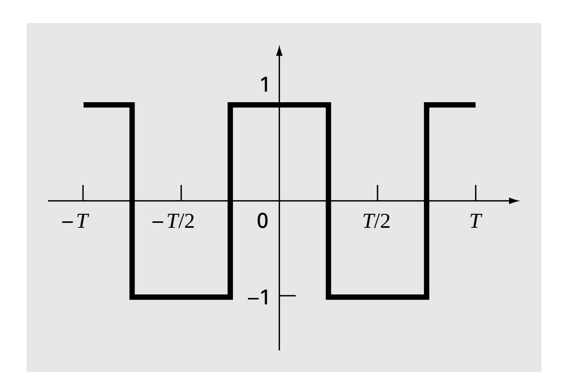


Figura 19.5
Una forma de onda cuadrada o rectangular con una altura de 2 y un periodo $T = 2\pi/\omega_0$.

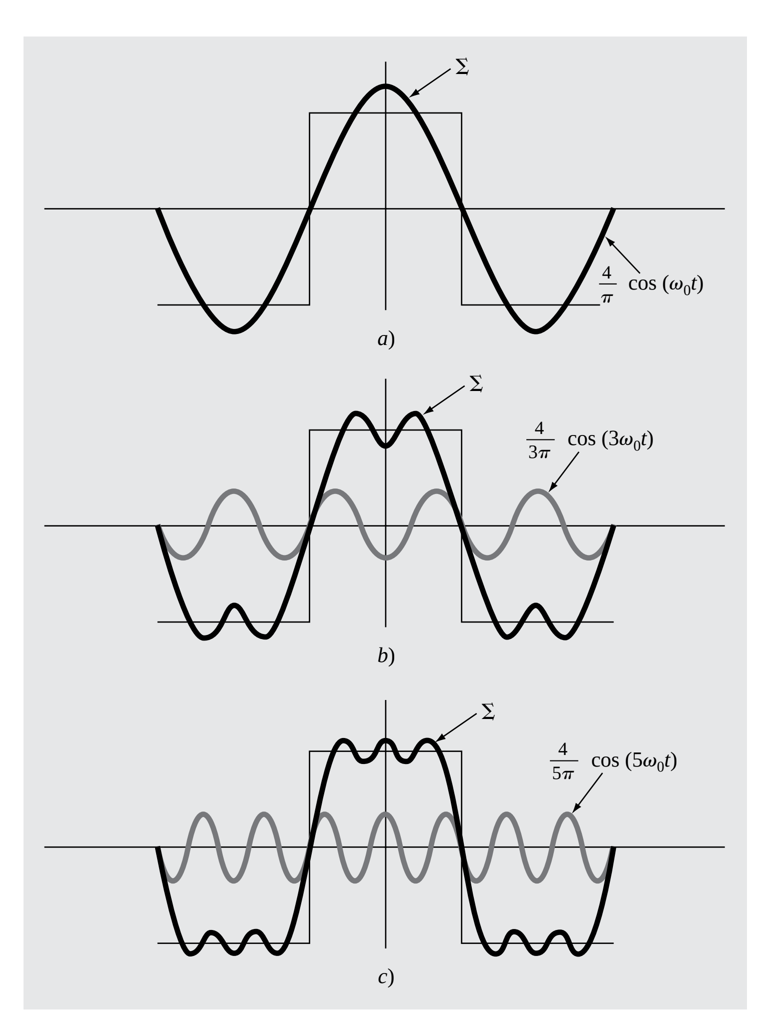


Figura 19.6
La aproximación de la serie de Fourier para la onda cuadrada de la figura 19.5. El conjunto de gráficas muestra la suma hasta, e incluyendo, el *a*) primer, *b*) segundo y *c*) tercer términos. Se presentan también los términos individuales que se fueron agregando en cada etapa.

$$a_k = \begin{cases} 4/(k\pi) & \text{para } k = 1, 5, 9, \ldots \\ -4/(k\pi) & \text{para } k = 3, 7, 11, \ldots \\ 0 & \text{para } k = \text{pares enteros} \end{cases}$$

De manera similar, se determina que todas las $b = 0$. Entonces, la aproximación de la serie de Fourier es

$$f(t) = \frac{4}{\pi}\cos(\omega_0 t) - \frac{4}{3\pi}\cos(3\omega_0 t) + \frac{4}{5\pi}\cos(5\omega_0 t) - \frac{4}{7\pi}\cos(7\omega_0 t) + \cdots$$

Los resultados hasta los primeros tres términos se muestran en la figura 19.6.

Debe mencionarse que a la onda cuadrada de la figura 19.5 se le llama *función par,* ya que $f(t) = f(-t)$. Otro ejemplo de una función par es $\cos(t)$. Se puede demostrar (Van Valkenburg, 1974) que las b_i en la serie de Fourier siempre son iguales a cero en las funciones pares. Observe también que las *funciones impares* son aquellas en las que $f(t) = -f(-t)$. La función $\text{sen}(t)$ es una función impar. En este caso las a_i serán iguales a cero.

Además de la forma trigonométrica de la ecuación (19.17), la serie de Fourier se expresa también en términos de funciones exponenciales como sigue (véase el cuadro 19.2 y el apéndice A)

$$f(t) = \sum_{k=-\infty}^{\infty} \tilde{c}_k e^{ik\omega_0 t} \tag{19.21}$$

donde $i = \sqrt{-1}$ y

$$\tilde{c}_k = \frac{1}{T}\int_{-T/2}^{T/2} f(t)e^{-ik\omega_0 t}\, dt \tag{19.22}$$

Esta fórmula alternativa tendrá utilidad a lo largo de lo que resta del capítulo.

Cuadro 19.2 Forma compleja de las series de Fourier

La forma trigonométrica de la serie de Fourier continua es

$$f(t) = a_0 + \sum_{k=1}^{\infty} [a_k \cos(k\omega_0 t) + b_k \,\text{sen}(k\omega_0 t)] \tag{C19.2.1}$$

A partir de la identidad de Euler, el seno y el coseno se expresan en forma exponencial como

$$\text{sen}\, x = \frac{e^{ix} - e^{-ix}}{2i} \tag{C19.2.2}$$

$$\cos x = \frac{e^{ix} + e^{-ix}}{2} \tag{C19.2.3}$$

las cuales se sustituyen en la ecuación (C19.2.1) para dar

$$f(t) = a_0 + \sum_{k=1}^{\infty} \left(e^{ik\omega_0 t} \frac{a_k - ib_k}{2} + e^{-ik\omega_0 t} \frac{a_k + ib_k}{2} \right) \tag{C19.2.4}$$

ya que $1/i = -i$. Podemos definir un conjunto de constantes

$$\tilde{c}_0 = a_0$$
$$\tilde{c}_k = \frac{a_k - ib_k}{2}$$
$$\tilde{c}_{-k} = \frac{a_{-k} - ib_{-k}}{2} = \frac{a_k + ib_k}{2} \tag{C19.2.5}$$

donde, debido a las propiedades de simetría del coseno y del seno, $a_k = a_{-k}$ y $b_k = -b_{-k}$. La ecuación (C19.2.4) puede, por lo tanto, reexpresarse como

$$f(t) = \tilde{c}_0 + \sum_{k=0}^{\infty} \tilde{c}_k e^{ik\omega_0 t} + \sum_{k=1}^{\infty} \tilde{c}_{-k} e^{-ik\omega_0 t}$$

o

$$f(t) = \sum_{k=0}^{\infty} \tilde{c}_k e^{ik\omega_0 t} + \sum_{k=1}^{\infty} \tilde{c}_{-k} e^{-ik\omega_0 t}$$

Para simplificar aún más, en lugar de sumar la segunda serie desde 1 hasta ∞, se realiza la suma de –1 a ∞,

$$f(t) = \sum_{k=0}^{\infty} \tilde{c}_k e^{ik\omega_0 t} + \sum_{k=-1}^{-\infty} \tilde{c}_k e^{ik\omega_0 t}$$

o

$$f(t) = \sum_{k=-\infty}^{\infty} \tilde{c}_k e^{ik\omega_0 t} \tag{C19.2.6}$$

donde la sumatoria incluye un término para $k = 0$.

Para evaluar las $\tilde{c}_k$, las ecuaciones (19.18) y (19.19) se sustituyen en la ecuación (C19.2.5) para obtener

$$\tilde{c}_k = \frac{1}{T}\int_{-T/2}^{T/2} f(t)\cos(k\omega_0 t)\, dt - i\frac{1}{T}\int_{-T/2}^{T/2} f(t)\,\text{sen}(k\omega_0 t)\, dt$$

Mediante las ecuaciones (C19.2.2) y (C19.2.3) y simplificando se obtiene

$$\tilde{c}_k = \frac{1}{T}\int_{-T/2}^{T/2} f(t)e^{-ik\omega_0 t}\, dt \tag{C19.2.7}$$

Por lo tanto, las ecuaciones (C19.2.6) y (C19.2.7) son las versiones complejas de las ecuaciones (19.17) a (19.20). Observe que el apéndice A incluye un resumen de las interrelaciones entre todas las formas de la serie de Fourier que se presentan en este capítulo.

19.3 DOMINIOS DE LA FRECUENCIA Y DEL TIEMPO

Hasta aquí, nuestro análisis de la aproximación de Fourier se ha limitado al *dominio del tiempo*. Esto se debe a que para la mayoría de nosotros resulta fácil conceptualizar el comportamiento de una función en esta dimensión. Aunque no sea muy familiar, el *dominio de la frecuencia* ofrece una perspectiva alternativa para caracterizar el comportamiento de funciones oscilantes.

Así, justo como se grafica la amplitud contra tiempo, de igual manera se grafica contra la frecuencia. Ambos tipos de expresión se ilustran en la figura 19.7*a*, donde se dibuja una gráfica en tres dimensiones de una función sinusoidal,

$$f(t) = C_1 \cos\left(t + \frac{\pi}{2}\right)$$

En esta gráfica, la magnitud o la amplitud de la curva, $f(t)$, es la variable dependiente; y las variables independientes son el tiempo t y la frecuencia $f = \omega_0/2\pi$. Así, los ejes de la amplitud y del tiempo forman un *plano de tiempo*; y los ejes amplitud y frecuencia, un *plano de frecuencia*. Por lo tanto, la sinusoide se concibe como si existiera a una distancia $1/T$ hacia afuera y a lo largo del eje de la frecuencia, y corriendo paralela a los ejes del tiempo. En consecuencia, cuando se habla acerca del comportamiento de la sinusoide en el dominio del tiempo, significa la proyección de la curva en el plano del tiempo (figura 19.7*b*). De manera similar, el comportamiento en el dominio de la frecuencia es tan sólo su proyección en el plano de la frecuencia.

Como se observa en la figura 19.7*c*, esta proyección es una medida de la amplitud positiva máxima de la sinusoide C_1. La oscilación completa de pico a pico es innecesaria debido a la simetría. Junto con la ubicación $1/T$ a lo largo del eje de la frecuencia, la figura 19.7*c* define ahora la amplitud y frecuencia de la sinusoide. Esta información es suficiente para reproducir la forma y el tamaño de la curva en el dominio del tiempo. Sin embargo, se requiere un parámetro más, el ángulo de fase, para ubicar la curva en relación con $t = 0$. En consecuencia, se debe incluir también un diagrama de fase, como el que se muestra en la figura 19.7*d*. El ángulo de fase se determina como la distancia (en radianes) desde cero al punto donde se presenta el pico positivo. Si el pico se presenta después del cero, se dice que está retrasada (recuerde nuestro análisis de retrasos y adelantos de la sección 19.1) y, por convención, al ángulo de fase se le antepone signo negativo. En forma opuesta, con un pico antes de cero se dice que está adelantada y el ángulo de fase es positivo. Así, en la figura 19.7, el pico está antes del cero y el ángulo de fase se grafica como $+\pi/2$. En la figura 19.8 se ilustran otras posibilidades.

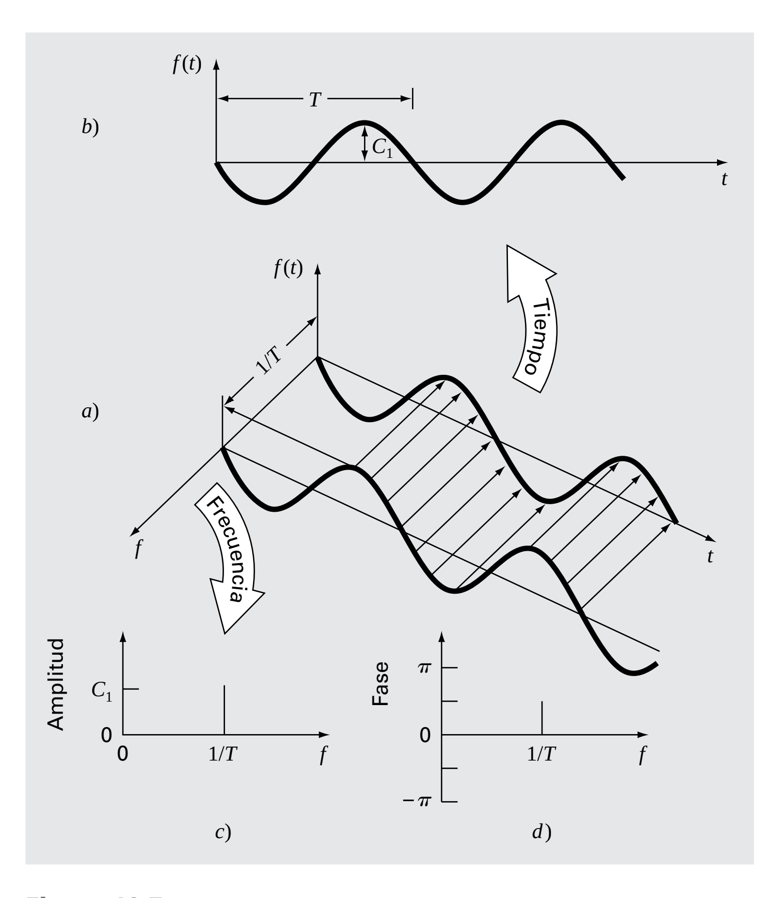


Figura 19.7
a) Una ilustración de cómo se representa una sinusoide en los dominios de la frecuencia y del tiempo. Se reproduce la proyección en el tiempo en *b*); mientras que la proyección de amplitud-frecuencia se reproduce en *c*). La proyección de fase-frecuencia se muestra en *d*).

Se puede observar ahora que las figuras 19.7*c* y 19.7*d* proporcionan una forma alternativa de presentar o resumir las características de la sinusoide de la figura 19.7*a*. Se hace referencia a ellas como *espectros de línea*. Se acepta que para una sola sinusoide estas líneas no son muy interesantes. Sin embargo, cuando se aplican a una situación más complicada, digamos, una serie de Fourier, se revela su poder y su valor. Por ejemplo, la figura 19.9 muestra el espectro de amplitud y el espectro de fase para la función de onda cuadrada del ejemplo 19.2.

Tales espectros ofrecen información que no aparece en el dominio del tiempo. Esto se puede ver al comparar las figuras 19.6 y 19.9. La figura 19.6 presenta dos perspectivas alternativas en el dominio del tiempo. La primera, la onda cuadrada original, no nos indica nada acerca de las sinusoides que comprende. La alternativa consiste en desplegar estas sinusoides [es decir, $(4/\pi)\cos(\omega_0 t)$, $-(4/3\pi)\cos(3\omega_0 t)$, $(4/5\pi)\cos(5\omega_0 t)$], etc. Esta alternativa no proporciona una visualización adecuada de la estructura de estas armónicas. Al contrario, las figuras 19.9*a* y 19.9*b* ofrecen una representación gráfica de esta estructura. Como tal, el espectro de línea representa "huellas dactilares" que nos pueden ayudar a caracterizar y entender una forma de onda complicada. En particular ellos son valiosos en casos no idealizados donde algunas veces nos permiten discernir una estructura, mientras que de otra manera obtendríamos sólo señales oscuras. En la siguiente sección se describirá la transformada de Fourier que nos permitirá extender tal análisis a ondas de forma no periódica.

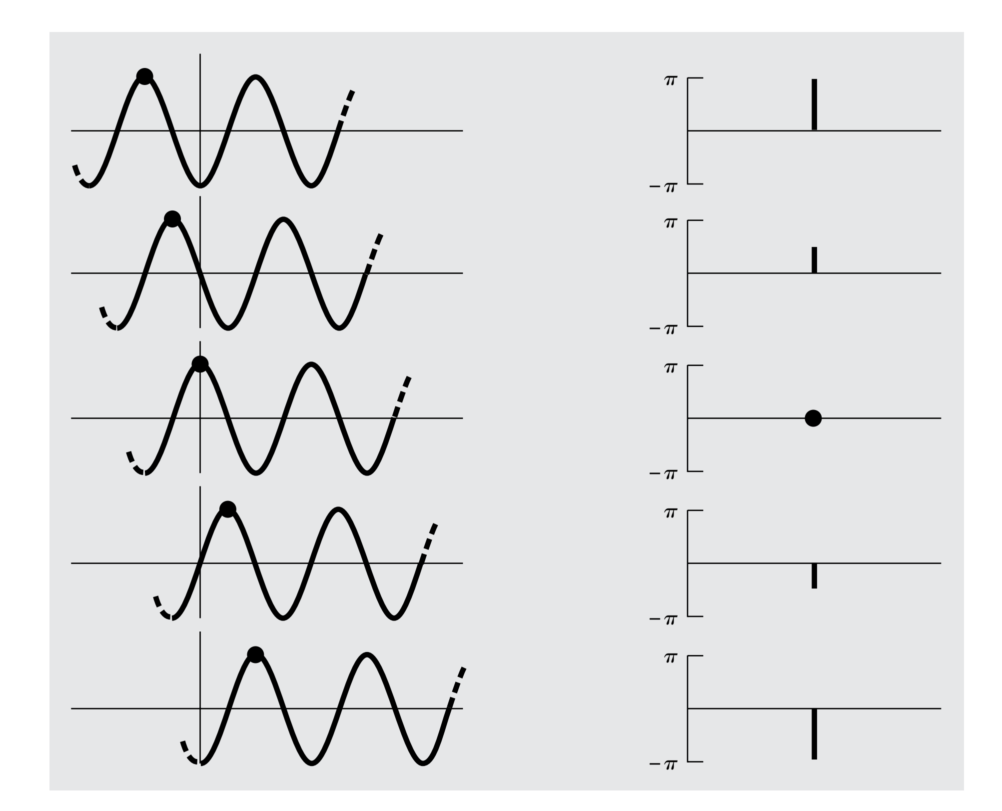


Figura 19.8
Varias fases de una sinusoide que muestran el espectro de fase correspondiente.

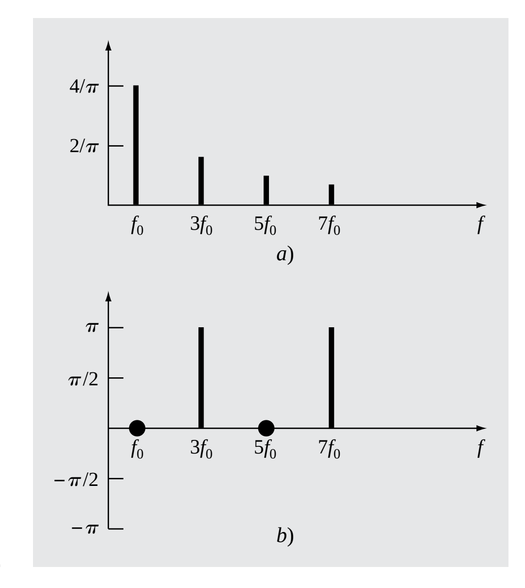


Figura 19.9
a) Espectro de amplitud y
b) espectro de fase para la onda cuadrada de la figura 19.5.

19.4 INTEGRAL Y TRANSFORMADA DE FOURIER

Aunque la serie de Fourier es una herramienta útil para investigar el espectro de una función periódica, existen muchas formas de onda que no se autorrepiten de manera regular. Por ejemplo, un relámpago ocurre sólo una vez (o al menos pasará mucho tiempo para que ocurra de nuevo); pero causará interferencia en los receptores que están operando en un amplio rango de frecuencias (por ejemplo, en televisores, radios, receptores de onda corta, etc.). Tal evidencia sugiere que una señal no recurrente como la producida por un relámpago exhibe un espectro de frecuencia continuo. Ya que fenómenos como éstos son de gran interés para los ingenieros, una alternativa a la serie de Fourier sería valiosa para analizar dichas formas de onda no periódicas.

La *integral de Fourier* es la principal herramienta para este propósito. Se puede obtener de la forma exponencial de la serie de Fourier

$$f(t) = \sum_{k=-\infty}^{\infty} \tilde{c}_k e^{ik\omega_0 t} \tag{19.23}$$

donde

$$\tilde{c}_k = \frac{1}{T}\int_{-T/2}^{T/2} f(t)e^{-ik\omega_0 t}\, dt \tag{19.24}$$

donde $\omega_0 = 2\pi/T$ y $k = 0, 1, 2,\ldots$

La transición de una función periódica a una no periódica se efectúa al permitir que el periodo tienda al infinito. En otras palabras, conforme T se vuelve infinito, la función nunca se repite y, de esta forma, se vuelve no periódica. Si se permite que ocurra esto, se puede demostrar (por ejemplo, Van Valkenburg, 1974; Hayt y Kemmerly, 1986) que la serie de Fourier se reduce a

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(i\omega_0)e^{i\omega_0 t}\, d\omega_0 \tag{19.25}$$

y los coeficientes se convierten en una función continua de la variable frecuencia ω, teniéndose que

$$F(i\omega_0) = \int_{-\infty}^{\infty} f(t)e^{-i\omega_0 t}\, dt \tag{19.26}$$

La función $F(i\omega_0)$, definida por la ecuación (19.26), se llama *integral de Fourier* de $f(t)$. Entonces, las ecuaciones (19.25) y (19.26) se conocen como el *par de transformadas de Fourier.* Así, además de llamarse integral de Fourier, $F(i\omega_0)$ también se denomina *transformada de Fourier* de $f(t)$. De igual manera, $f(t)$, como se define en la ecuación (19.25), se conoce como *transformada inversa de Fourier* de $F(i\omega_0)$. Así, el par nos permite transformar entre uno y otro de los dominios de la frecuencia y del tiempo para una señal no periódica.

La diferencia entre la serie de Fourier y la transformada de Fourier ahora será clara. La principal diferencia radica en que cada una se aplica a un tipo diferente de funciones (las series a formas de onda periódicas y la transformada a las no periódicas). Además de esta diferencia principal, los dos procedimientos difieren en cómo se mueven entre los dominios de la frecuencia y del tiempo. La serie de Fourier convierte una función continua y periódica en el dominio del tiempo, a magnitudes de frecuencia discretas en el dominio de la frecuencia. Al contrario, la transformada de Fourier convierte una función continua en el dominio del tiempo en una función continua en el dominio de la frecuencia. De esta manera, el espectro de frecuencia discreto generado por la serie de Fourier es análogo a un espectro de frecuencia continuo generado por la transformada de Fourier.

El paso de un espectro continuo en uno discreto se puede ilustrar gráficamente. En la figura 19.10*a*, se observa el tren de pulsos de ondas rectangulares con amplitudes de pulsación iguales a la mitad del periodo, asociado con su correspondiente espectro discreto. Esta función es la misma que se investigó antes en el ejemplo 19.2, sólo que en este caso está corrida verticalmente.

En la figura 19.10*b*, al duplicar el periodo en el tren de pulsos se tienen dos efectos sobre el espectro. Primero, se agregan dos líneas de frecuencia a cada lado de las componentes originales. Segundo, se reducen las amplitudes de las componentes.

Conforme el periodo se aproxima al infinito, dichos efectos generan líneas espectrales cada vez más comprimidas, hasta que el espacio entre las líneas tiende a cero. En el límite, las series convergen a la integral de Fourier continua, como se muestra en la figura 19.10*c*.

Ahora que se ha presentado una forma para analizar una señal no periódica, veremos el paso final en nuestro desarrollo. En la siguiente sección analizaremos el hecho de que una señal rara vez está caracterizada como una función continua que se necesita para implementar la ecuación (19.26). En lugar de esto, los datos invariablemente están en forma discreta. Ahora se mostrará cómo calcular la transformada de Fourier a partir de mediciones discretas.

19.5 TRANSFORMADA DISCRETA DE FOURIER (TDF)

En ingeniería, las funciones en general se representan por conjuntos finitos de valores discretos. Es decir, los datos con frecuencia se obtienen de, o convierten a, una forma discreta. Como se indica en la figura 19.11, se puede dividir un intervalo de 0 a t en N subintervalos de igual tamaño $\Delta t = T/N$. El subíndice n se emplea para designar los tiempos discretos a los cuales se toman las muestras. Así, f_n designa un valor de la función continua $f(t)$ tomado en t_n.

Observe que los puntos asociados con datos se especifican en $n = 0, 1, 2, \ldots, N - 1$. No hay un valor en $n = N$. (Véase Ramírez, 1985, para la razón de la exclusión de f_N.)

Para el sistema de la figura 19.11 se escribe la transformada discreta de Fourier como

$$F_k = \sum_{n=0}^{N-1} f_n e^{-ik\omega_0 n} \quad \text{para } k = 0 \text{ a } N-1 \tag{19.27}$$

y la transformada inversa de Fourier como

$$f_n = \frac{1}{N}\sum_{k=0}^{N-1} F_k e^{ik\omega_0 n} \quad \text{para } n = 0 \text{ a } N-1 \tag{19.28}$$

donde $\omega_0 = 2\pi/N$.

Las ecuaciones (19.27) y (19.28) representan las análogas discretas de las ecuaciones (19.26) y (19.25), respectivamente. Como tales, ellas se emplean para calcular tanto la transformada directa

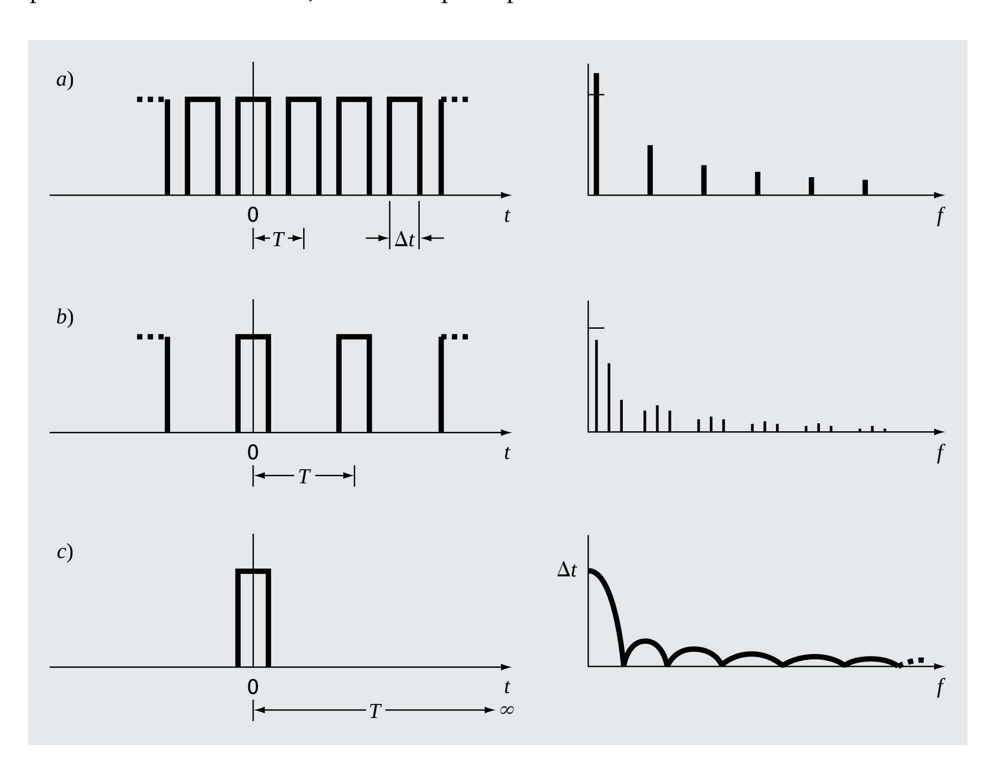


Figura 19.10
Ilustración de cómo el espectro de frecuencia discreta de una serie de Fourier para un tren de pulsos *a*) se aproxima a un espectro de frecuencia continua de una integral de Fourier *c*) conforme el periodo se aproxima al infinito.

como la inversa de Fourier, para datos discretos. Aunque es posible realizar tales cálculos a mano, son bastante laboriosos. Como lo expresa la ecuación (19.27), la TDF requiere N^2 operaciones complejas. Así, es necesario desarrollar un algoritmo computacional para implementar la TDF.

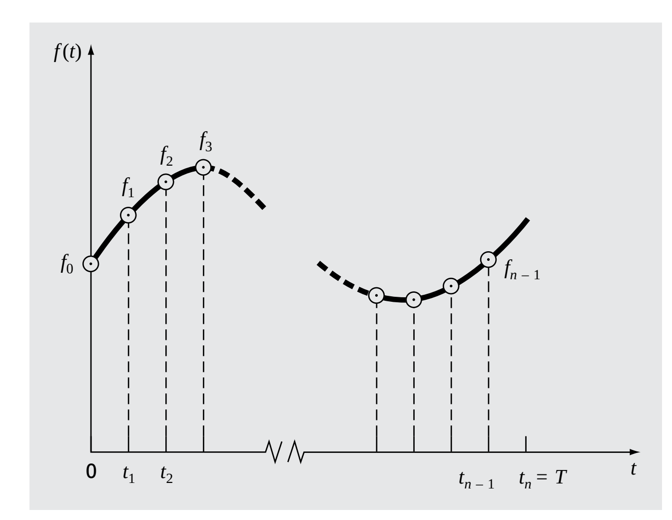


Figura 19.11
Los puntos muestrales de la serie discreta de Fourier.

Algoritmo computacional para la TDF Observe que el factor $1/N$ en la ecuación (19.28) es sólo un factor de escala que se puede incluir tanto en la ecuación (19.27) como en la (19.28), pero no en ambas. En nuestro algoritmo computacional, lo incluiremos en la ecuación (19.27) para que el primer coeficiente F_0 (que es el análogo del coeficiente continuo a_0) sea igual a la media aritmética de las muestras. También, usaremos la identidad de Euler para implementar un algoritmo con lenguajes que no contengan datos de variables complejas,

$$e^{\pm ia} = \cos a \pm i \operatorname{sen} a$$

y después volver a expresar las ecuaciones (19.27) y (19.28) como

$$F_k = \frac{1}{N}\sum_{n=0}^{N} [\, f_n \cos(k\omega_0 n) - i f_n \operatorname{sen}(k\omega_0 n)] \tag{19.29}$$

y

$$f_n = \sum_{k=0}^{N-1} [F_k \cos(k\omega_0 n) + i F_k \operatorname{sen}(k\omega_0 n)] \tag{19.30}$$

El pseudocódigo para implementar la ecuación (19.29) se muestra en la figura 19.12. Este algoritmo se puede desarrollar como un programa computacional para calcular la TDF. Los resultados de tal programa se tienen en la figura 19.13 para el análisis de una función coseno.

```
DOFOR k = 0, N - 1
  DOFOR n = 0, N - 1
    angle = kω0n
    realk = realk + fn cos(angle)/N
    imaginaryk = imaginaryk - fn sin(angle)/N
  END DO
END DO
```

Figura 19.12
Pseudocódigo para el cálculo de la TDF.

INDICE	f(t)	REAL	IMAGINARIA
0	1.000	0.000	0.000
1	0.707	0.000	0.000
2	0.000	0.500	0.000
3	−0.707	0.000	0.000
4	−1.000	0.000	0.000
5	−0.707	0.000	0.000
6	0.000	0.000	0.000
7	0.707	0.000	0.000
8	1.000	0.000	0.000
9	0.707	0.000	0.000
10	0.000	0.000	0.000
11	−0.707	0.000	0.000
12	−1.000	0.000	0.000
13	−0.707	0.000	0.000
14	0.000	0.500	0.000
15	0.707	0.000	0.000

Figura 19.13
Resultados obtenidos con un programa basado en el algoritmo de la figura 19.12 para la TDF con los datos generados por una función coseno $f(t) = \cos[2\pi(12.5)t]$ en 32 puntos con $\Delta t = 0.01$ s.

19.6 TRANSFORMADA RÁPIDA DE FOURIER (TRF)

Aunque el algoritmo descrito en la sección anterior calcula de manera adecuada la TDF, es computacionalmente laborioso debido a que se requieren N^2 operaciones. En consecuencia, aun con muestras de un tamaño moderado, la determinación directa de la TDF llega a consumir mucho tiempo.

La *transformada rápida de Fourier*, o TRF, es un algoritmo que se desarrolló para calcular la TDF en una forma extremadamente económica. Su velocidad proviene del hecho de que utiliza los resultados de cálculos previos para reducir el número de operaciones. En particular, aprovecha la periodicidad y simetría de las funciones trigonométricas para calcular la transformada con aproximadamente $N \log_2 N$ operaciones (véase figura 19.14). Así, para $N = 50$ muestras, la TRF es cerca de 10 veces más rápida que la TDF estándar. Para $N = 1\,000$, es alrededor de 100 veces más rápida.

El primer algoritmo para la TRF fue desarrollado por Gauss a principios del siglo XIX (Heideman y cols., 1984). Otras contribuciones importantes fueron hechas por Runge, Danielson, Lanczos y otros a comienzos del siglo XX. Sin embargo, como calcular manualmente las transformadas discretas tomaba días o semanas, no atraían mucho el interés antes del desarrollo de la moderna computadora digital.

En 1965, J. W. Cooley y J. W. Tukey publicaron un artículo clave, en el cual se propuso un algoritmo para el cálculo de la TRF. Dicho esquema, similar a aquel de Gauss y de otros investigadores anteriores, se conoce como *algoritmo de Cooley-Tukey*. En la actualidad, existen otros procedimientos que son adaptaciones de este método.

La idea básica detrás de cada uno de estos algoritmos es que una TDF de longitud N se descompone, o "particiona" sucesivamente en TDF más pequeñas. Hay una variedad de formas diferentes de aplicar este principio. Por ejemplo, el algoritmo de Cooley-Tukey usa las llamadas técnicas de *partición en el tiempo.* En esta sección se describirá un procedimiento alternativo llamado *algoritmo de Sande-Tukey*. Este método pertenece a otra clase de algoritmos que se denominan técnicas de *partición en frecuencia*. La distinción entre las dos clases se analizará tras desarrollar el método.

19.6.1 Algoritmo de Sande-Tukey

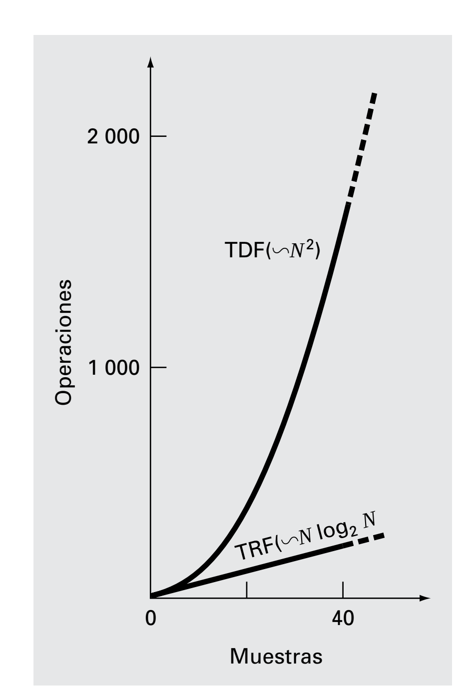


Figura 19.14
Gráfica del número de operaciones contra tamaño de la muestra de la TDF estándar y la TRF.

En el presente caso, se supondrá que N es una potencia entera de 2,

$$N = 2^M \tag{19.31}$$

donde M es un entero. Se introduce esta restricción para simplificar el algoritmo resultante. Ahora, recuerde que la TDF se puede representar de manera general como

$$F_k = \sum_{k=0}^{N-1} f_n e^{-i(2\pi/N)nk} \qquad \text{para } k = 0 \text{ a } N-1 \tag{19.32}$$

donde $2\pi/N = \omega_0$. La ecuación (19.32) se expresa también como

$$F_k = \sum_{n=0}^{N-1} f_n W^{nk}$$

donde W es una función ponderada de valor complejo definida como

$$W = e^{-i(2\pi/N)} \tag{19.33}$$

Suponga ahora que la muestra se divide a la mitad y la ecuación (19.32) se expresa en términos de los primeros y últimos $N/2$ puntos:

$$F_k = \sum_{n=0}^{(N/2)-1} f_n e^{-i(2\pi/N)kn} + \sum_{n=N/2}^{N-1} f_n e^{-i(2\pi/N)kn}$$

donde $k = 0, 1, 2,\ldots, N - 1$. Se crea una nueva variable, $m = n - N/2$, para que los límites de la segunda sumatoria sean consistentes con la primera,

$$F_k = \sum_{n=0}^{(N/2)-1} f_n e^{-i(2\pi/N)kn} + \sum_{m=0}^{(N/2)-1} f_{m+N/2} e^{-i(2\pi/N)k(m+N/2)}$$

o

$$F_k = \sum_{n=0}^{(N/2)-1} (f_n + e^{-i\pi k} f_{n+N/2}) e^{-i2\pi kn/N} \tag{19.34}$$

Ahora, advierta que el factor $e^{-i\pi k} = (-1)^k$. De esta forma, para puntos pares es igual a 1 y para los impares es igual a –1. Por lo tanto, el siguiente paso en el método consiste en separar la ecuación (19.34) de acuerdo con valores pares o impares de k. Para los valores pares,

$$F_{2k} = \sum_{n=0}^{(N/2)-1} (f_n + f_{n+N/2}) e^{-i2\pi(2k)n/N} = \sum_{n=0}^{(N/2)-1} (f_n + f_{n+N/2}) e^{-i2\pi kn/(N/2)}$$

y para los valores impares,

$$F_{2k+1} = \sum_{n=0}^{(N/2)-1} (f_n - f_{n+N/2}) e^{-i2\pi(2k+1)n/N}$$
$$= \sum_{n=0}^{(N/2)-1} (f_n - f_{n+N/2}) e^{-i2\pi n/N} e^{-i2\pi kn/(N/2)}$$

para $k = 0, 1, 2, \ldots, (N/2) - 1$.

Estas ecuaciones se expresan también en términos de la ecuación (19.33). Para los valores pares,

$$F_{2k} = \sum_{n=0}^{(N/2)-1} (f_n + f_{n+N/2}) W^{2kn}$$

y para los valores impares,

$$F_{2k+1} = \sum_{n=0}^{(N/2)-1} (f_n - f_{n+N/2}) W^n W^{2kn}$$

Ahora, realizaremos una observación clave: esas expresiones pares e impares se pueden interpretar como si fueran iguales a las transformadas secuenciales de longitud $(N/2)$

$$g_n = f_n + f_{n+N/2} \tag{19.35}$$

y

$$h_n = (f_n - f_{n+N/2}) W^n \qquad \text{para } n = 0, 1, 2, \ldots, (N/2) - 1 \tag{19.36}$$

De esta manera, en forma directa resulta que

$$\left.\begin{aligned} F_{2k} &= G_k \\ F_{2k+1} &= H_k \end{aligned}\right\} \quad \text{para } k = 0, 1, 2, \ldots, (N/2) - 1$$

En otras palabras, se reemplazó un cálculo de N puntos por dos cálculos de $(N/2)$ puntos. Puesto que cada uno de los últimos requiere aproximadamente $(N/2)^2$ multiplicaciones y sumas complejas, el procedimiento permite un ahorro de un factor de 2 (es decir, N^2 contra $2(N/2)^2 = N^2/2$).

El esquema se ilustra en la figura 19.15 para $N = 8$. La TDF se calcula formando primero la secuencia g^n y h^n y calculando después las $N/2$ TDF para obtener las transformadas numeradas pares e impares. Algunas veces los pesos W^n se llaman *factores de giro.*

Ahora es claro que este procedimiento de “divide y vencerás” se puede repetir en la segunda etapa. Así, calculamos la TDF de $N/4$ puntos de las cuatro secuencias de $N/4$ compuestas de los primeros y últimos $N/4$ puntos de las ecuaciones (19.35) y (19.36).

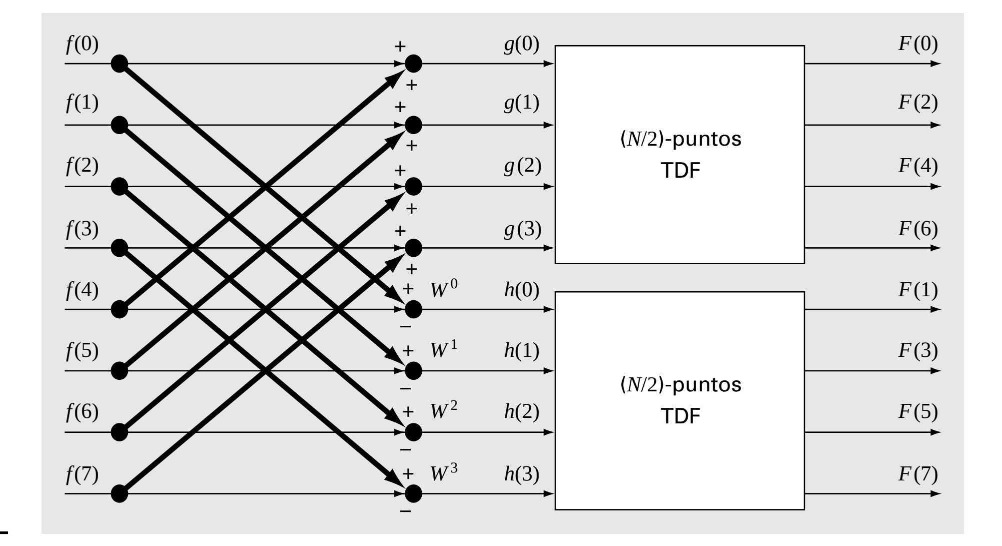


Figura 19.15
Diagrama de flujo de la primera etapa en una descomposición por partición en frecuencia de una TDF con *N* puntos en dos TDF con (*N*/2) puntos para *N* = 8.

Se continúa la estrategia hasta su inevitable conclusión, cuando *N*/2 de TDF de dos puntos se hayan calculado (figura 19.16). El número total de cálculos para el cálculo completo es del orden de $N \log_2 N$. La diferencia entre este nivel de esfuerzo y el de la TDF estándar (figura 19.14) ilustra por qué es tan importante la TRF.

Algoritmo computacional Es relativamente sencillo expresar la figura 19.16 como un algoritmo. Como en el caso del algoritmo para la TDF de la figura 19.12, se usará la identidad de Euler,

$$e^{\pm ia} = \cos a \pm i \text{ sen } a$$

para implementar el algoritmo en lenguajes que no emplean en forma explícita variables complejas.

Una inspección cercana a la figura 19.16 indica que su molécula computacional fundamental es la llamada *red mariposa*, ilustrada en la figura 19.17*a*. El pseudocódigo para implementar una de esas moléculas se muestra en la figura 19.17*b*.

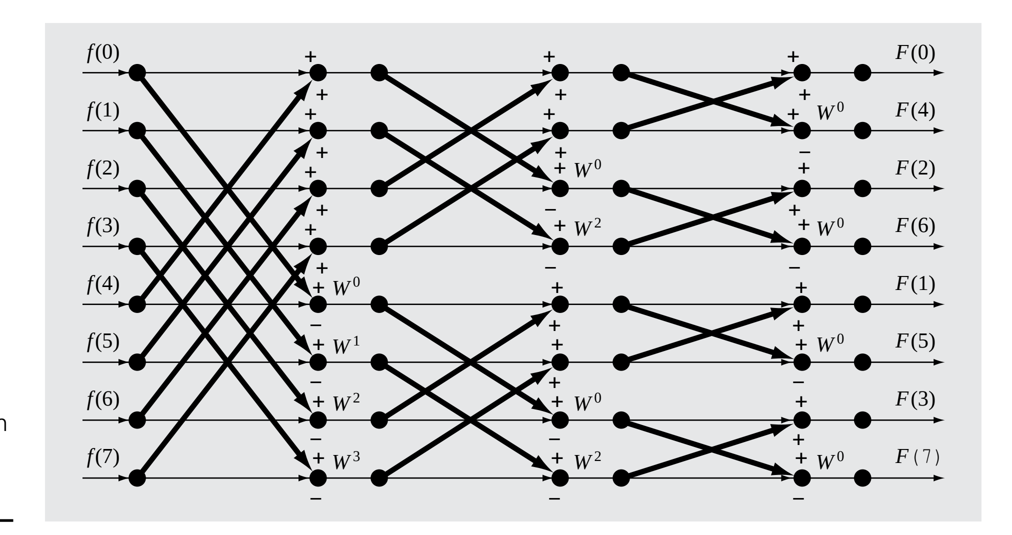


Figura 19.16
Diagrama de flujo de la descomposición completa por partición en frecuencia de una TDF con ocho puntos.

El pseudocódigo para la TRF se da en la figura 19.18. La primera parte consiste, en esencia, en tres ciclos anidados para implementar el cuerpo computacional de la figura 19.16. Observe que los datos reales se guardan originalmente en el arreglo *x*. También observe que el ciclo exterior pasa a través de las *M* etapas [recuerde la ecuación (19.31)] del diagrama de flujo.

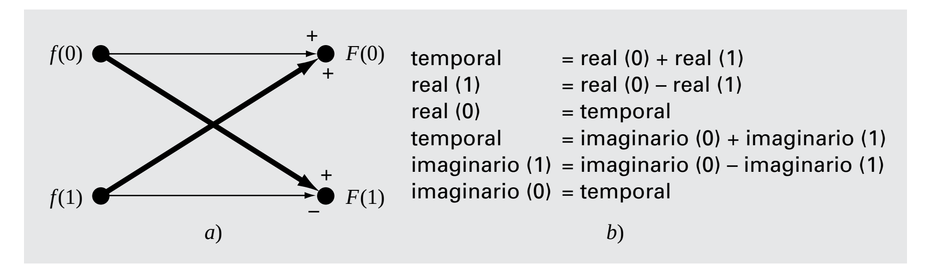


Figura 19.17
a) Una red mariposa que representa el cálculo fundamental de la figura 19.16. *b*) Pseudocódigo para implementar *a*).

```
a)
m = LOG(N)/LOG(2)
N2 = N
DOFOR k = 1, m
  N1 = N2
  N2 = N2/2
  angle = 0
  arg = 2π/N1
  DOFOR j = 0, N2 - 1
    c = cos(angle)
    s = -sin(angle)
    DOFOR i = j, N - 1, N1
      kk = i + N2
      xt = x(i) - x(kk)
      x(i) = x(i) + x(kk)
      yt = y(i) - y(kk)
      y(i) = y(i) + y(kk)
      x(kk) = xt * c - yt * s
      y(kk) = yt * c + xt * s
    END DO
    angle = (j + 1) * arg
  END DO
END DO
```

```
b)
j = 0
DOFOR i = 0, N - 2
  IF (i < J) THEN
      xt = xj
      xj = xi
      xi = xt
      yt = yj
      yj = yi
      yi = yt
   END IF
   k = N/2
   DO
     IF (k ≥ j + 1) EXIT
     j = j - k
     k = k/2
  END DO
  j = j + k
END DO
DOFOR i = 0, N - 1
  x(i) = x(i)/N
  y(i) = y(i)/N
END DO
```

Figura 19.18
Pseudocódigo para implementar una TRF con partición en frecuencia. Observe que el pseudocódigo está compuesto por dos partes: *a*) la TRF en sí y *b*) una rutina de inversión de bits para ordenar los coeficientes de Fourier resultantes.

Después de que se ejecuta esta primera parte, se habrán calculado las TDF, pero en desorden (véase el lado derecho de la figura 19.16). Es posible ordenar esos coeficientes de Fourier mediante un procedimiento llamado de *inversión del bit*. Si los subíndices 0 al 7 se expresan en forma binaria, se obtiene el orden correcto al invertir esos bits (figura 19.19). La segunda parte del algoritmo realiza este procedimiento.

19.6.2 Algoritmo de Cooley-Tukey

La figura 19.20 muestra una red de flujo para implementar el algoritmo de Cooley-Tukey. Para este caso, la muestra se divide inicialmente en puntos numerados pares e impares, y los resultados finales están en el orden correcto.

Este procedimiento se llama una *partición en el tiempo*. Es el inverso del algoritmo de Sande-Tukey descrito en la sección anterior. Aunque las dos clases de métodos difieren en organización, ambos presentan las $N \log_2 N$ operaciones que son la fortaleza del procedimiento de la TRF.

19.7 EL ESPECTRO DE POTENCIA

La TRF tiene diversas aplicaciones en ingeniería que van desde el análisis de vibración de estructuras y mecanismos, hasta el procesamiento de señales. Como se describió antes, el espectro de amplitud y fase proporciona un medio para entender la estructura de señales bastante aleatorias. De manera similar, un análisis útil llamado espectro de potencia se puede desarrollar a partir de la transformada de Fourier.

En desorden (decimal)		En desorden (binario)		En orden de bits invertidos (binario)		Resultado final (decimal)
F(0)		F(000)		F(000)		F(0)
F(4)		F(100)		F(001)		F(1)
F(2)		F(010)		F(010)		F(2)
F(6)	⇒	F(110)	⇒	F(011)	⇒	F(3)
F(1)		F(001)		F(100)		F(4)
F(5)		F(101)		F(101)		F(5)
F(3)		F(011)		F(110)		F(6)
F(7)		F(111)		F (111)		F(7)

Figura 19.19
Ilustración del proceso de inversión de bits.

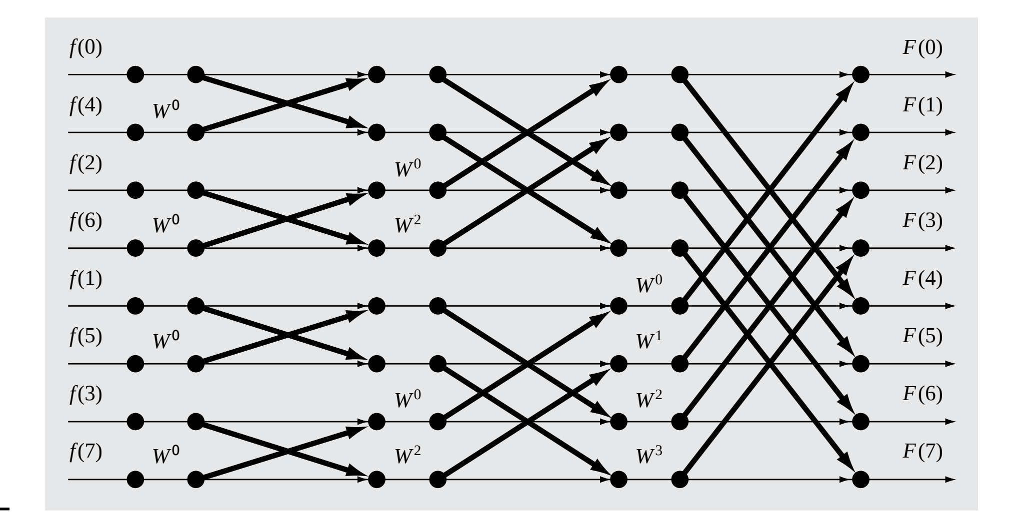


Figura 19.20
Diagrama de flujo de una TRF con partición en el tiempo para una TDF de 8 puntos.

Como su nombre indica, el espectro de potencia se obtiene del análisis de la potencia de salida en sistemas eléctricos. En términos matemáticos, la potencia de una señal periódica en el dominio del tiempo se define como

$$P = \frac{1}{T}\int_{-T/2}^{T/2} f^2(t)\, dt \tag{19.37}$$

Ahora, otra forma de entender esta información es expresándola en el dominio de la frecuencia y calculando la potencia asociada a cada componente de frecuencia. Después esta información se despliega como un *espectro de potencia*, es decir, una gráfica de la potencia contra la frecuencia.

Si la serie de Fourier para $f(t)$ es

$$f(t) = \sum_{k=-\infty}^{\infty} F_k e^{ik\omega_0 t} \tag{19.38}$$

se satisface la siguiente relación (véase Gabel y Roberts, 1987, para más detalles):

$$\frac{1}{T}\int_{-T/2}^{T/2} f^2(t)\, dt = \sum_{k=-\infty}^{\infty} |F_k|^2 \tag{19.39}$$

De esta forma, la potencia en $f(t)$ se determina al sumar los cuadrados de los coeficientes de Fourier, es decir, las potencias asociadas con los componentes de frecuencia individual.

Ahora, recuerde que, en esta representación, la armónica real simple consta de ambos componentes de frecuencia en $\pm k\omega_0$. También sabemos que los coeficientes positivos y negativos son iguales. Por lo tanto, la potencia en $f_k(t)$, la k-ésima armónica real de $f(t)$, es

$$p_k = 2\,|F_k|^2 \tag{19.40}$$

El espectro de potencia es la gráfica de p_k en función de la frecuencia $k\omega_0$. Dedicaremos la sección 20.3 a una aplicación en ingeniería que emplea la TRF y el espectro de potencia obtenido por medio de un paquete de software.

Información adicional Lo anterior ha sido una breve introducción a la aproximación de Fourier y a la TRF. Se puede encontrar información adicional sobre la primera en Van Valkenburg (1974), Chirlian (1969), y Hayt y Kemmerly (1986). Las referencias sobre la TRF se encuentran en Davis y Rabinowitz (1975); Cooley, Lewis y Welch (1977), y Brigham (1974). Buenas introducciones a ambos temas se encuentran en Ramírez (1985), Oppenheim y Schafer (1975), Gabel y Roberts (1987).

19.8 AJUSTE DE CURVAS CON PAQUETES DE SOFTWARE

Los paquetes de software tienen grandes posibilidades para el ajuste de curvas. En esta sección daremos una muestra de las más usuales.

19.8.1 Excel

En el presente contexto, la aplicación más útil de Excel es en el análisis de regresión y, en menor extensión, en la interpolación polinomial. Además de algunas funciones incorporadas (véase la tabla 19.l), existen dos formas principales en las que se puede emplear esta posibilidad: el comando Trendline y el Data Analysis Toolpack (paquete de herramientas para el análisis de datos).

Tabla 19.1 Funciones de Excel interconstruidas relacionadas con el ajuste de datos por regresión.

Función	Descripción	Función	Descripción
FORECAST	Da un valor junto con una tendencia lineal	LOGEST	Da los parámetros de una tendencia exponencial
GROWTH	Da valores junto con una tendencia exponencial	SLOPE	Da la pendiente de la recta de regresión lineal
INTERCEPT	Da la intersección de la recta de regresión lineal	TREND	Da valores junto con una tendencia lineal
LINEST	Da los parámetros de una tendencia lineal		

El comando Trendline (menú Insert) Este comando permite agregar varios modelos de tendencia a una gráfica. Tales modelos comprenden ajustes lineales, polinomiales, logarítmicos, exponenciales, de potencia y de promedio móviles. El siguiente ejemplo ilustra cómo utilizar el comando Trendline.

EJEMPLO 19.3 Uso del comando Trendline de Excel

Planteamiento del problema Usted habrá notado que varios de los ajustes que tiene Trendline fueron analizados ya en el capítulo 17 (por ejemplo, lineal, polinomial, exponencial y de potencia). Una posibilidad adicional es el modelo logarítmico

$$y = a_0 + a_1 \log x$$

Ajuste los siguientes datos con este modelo usando el comando Trendline de Excel:

x	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5
y	0.53	0.69	1.5	1.5	2	2.06	2.28	2.23	2.73	2.42	2.79

Solución Para usar el comando Trendline, se debe crear una gráfica que relacione una serie de variables dependientes y de variables independientes. Para el caso actual, se crea primero una gráfica XY de los datos.

SOFTWARE

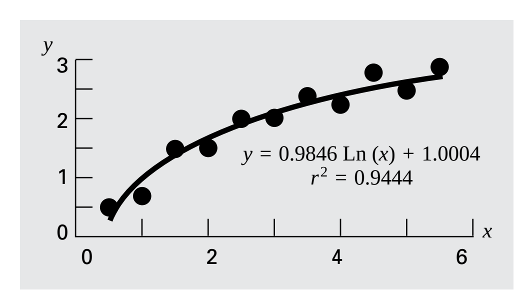


Figura 19.21
Ajuste de un modelo logarítmico a los datos del ejemplo 19.3.

En seguida, se puede seleccionar la gráfica (haciendo clic sobre ella) y la serie (posicionando la flecha del mouse en uno de los valores y haciendo clic derecho). Aparecerá un menú donde usted puede seleccionar **Add Trendline** ("Agregar línea de tendencia").

Entonces se abre una caja para formatear la línea de tendencia (Format Trendline), donde puede seleccionar el Trend/RegressionType (tipo de tendencia/regresión) como Logarithmic (logarítmico). Además, seleccione **Display Equation on chart** y **Display R-squared value on chart**. Cuando se cierre la caja, aparece la curva de ajuste resultante junto con r^2 como se muestra en la figura 19.21.

El comando Trendline proporciona una manera fácil para ajustar a los datos varios modelos que se usan comúnmente. Además, la opción **Polinomial** se incluye también para que se pueda usar la interpolación polinomial. Sin embargo, como su contenido estadístico está limitado a r^2, esto significa que no permite obtener gráficas de inferencias estadísticas respecto al ajuste del modelo. El paquete de herramientas para el análisis de datos (Data Analysis Toolpack) que se describirá a continuación ofrece una excelente alternativa en casos donde son necesarias las inferencias.

El paquete de herramientas para el análisis de datos (Data Analysis Toolpack) Este paquete adicional de Excel tiene amplias posibilidades para el ajuste de curvas por mínimos cuadrados lineales generales. Como se describió en la sección 17.4, tales modelos son de la forma general

$$y = a_0z_0 + a_1z_1 + a_2z_2 + \cdots + a_mz_m + e \tag{17.23}$$

donde $z_0, z_1, \ldots, z_m$ son $m + 1$ funciones diferentes. El siguiente ejemplo ilustra cómo tales modelos se pueden ajustar con Excel.

EJEMPLO 19.4 Uso del paquete de herramientas para el análisis de datos (Data Analysis Toolpack) de Excel

Planteamiento del problema Los siguientes datos son la pendiente, el radio hidráulico y la velocidad del agua que fluye en un canal:

S, m/m	0.0002	0.0002	0.0005	0.0005	0.001	0.001
R, m	0.2	0.5	0.2	0.5	0.2	0.5
U, m/s	0.25	0.5	0.4	0.75	0.5	1

Se tienen razones teóricas para creer que los datos se pueden ajustar a un modelo de potencias de la forma

$$U = \alpha S^{\sigma} R^{\rho}$$

donde α, σ y ρ son coeficientes obtenidos de manera empírica. Existen razones teóricas para creer que σ y ρ serán aproximadamente de 0.5 y 0.667, respectivamente. Ajuste estos datos con Excel y determine si los valores estimados con la regresión contradicen los valores esperados de los coeficientes del modelo.

Solución En el modelo de potencias se aplican primero logaritmos para convertirlo a la forma lineal de la ecuación (17.23),

$U = \log \alpha + \sigma \log S + \rho \log R$

Se puede desarrollar una hoja de cálculo en Excel, tanto con los datos originales como con sus respectivos logaritmos, como en la siguiente tabla:

	A	B	C	D	E	F
1	S	R	U	log (S)	log (R)	log (U)
2	0.0002	0.2	0.25	**–3.69897**	–0.69897	–0.60206
3	0.0002	0.5	0.5	–3.69897	–0.30103	–0.30103
4	0.0005	0.2	0.4	–3.30103	–0.69897	–0.39794
5	0.0005	0.5	0.75	–3.30103	–0.30103	–0.12494
6	0.001	0.2	0.5	–3	–0.69897	–0.30103
7	0.001	0.5	1	–3	–0.30103	0

=log(A2)

Como se indica, una manera eficiente para generar los logaritmos es tecleando la fórmula para calcular el primer log(S). Después se copia esta fórmula a la derecha y se baja para generar los otros logaritmos.

Debido a su estatus como agregado a la versión de Excel disponible en el momento de la edición en inglés de este libro, algunas veces hay que cargar en Excel el paquete de herramientas para el análisis de datos. Para hacer esto, elija File, Options, Add-Ins, Manage Excel Add-ins. Luego marque la caja Analysis ToolPak. Entonces debe aparecer como instalado el ToolPak y un botón para acceder a éste en su pestaña "Data".

Después de seleccionar **Data Analysis** del menú Data, aparecerá en pantalla un menú de Data Analysis que contiene un gran número de rutinas orientadas estadísticamente. Seleccione **Regression** y se desplegará un cuadro de diálogo que esperará que se le proporcione información sobre la regresión. Después de estar seguros que se ha seleccionado la instrucción por default **New Worksheet Ply**, dé F2:F7 como el rango y y D2:E7 como el rango x, y seleccione OK. Se creará la siguiente hoja de cálculo:

	A	B	C	D	E	F	G
1	RESUMEN DE RESULTADOS						
2							
3	*Estadística de regresión*						
4	Múltiple R	0.998353					
5	R cuadrada	0.996708					
6	Aj. de R cuadrada	0.994513					
7	Error estándar	0.015559					
8	Observaciones	6					
9							
10	ANOVA						
11		*df*	*SS*	*MS*	*F*	*Significancia F*	
12	Regresión	2	0.219867	0.10993	454.1106	0.0001889	
13	Residual	3	0.000726	0.00024			
14	Total	5	0.220593				
15							
16		*Coeficientes*	*Error estándar*	*Estad. t*	*Valor P*	*Inf. al 95%*	*Sup. al 95%*
17	Intersección	**1.522452**	0.075932	20.05010	0.000271	1.2808009	1.7641028
18	X Variable 1	**0.433137**	0.022189	19.52030	0.000294	**0.362521**	**0.503752**
19	X Variable 2	**0.732993**	0.031924	22.96038	0.000181	**0.631395**	**0.834590**

De esta manera, el ajuste resultante es

$\log U = 1.522 + 0.433 \log S + 0.733 \log R$

o tomando antilogaritmos,

$U = 33.3S^{0.433}R^{0.733}$

SOFTWARE

Observe que se generaron intervalos de confianza de 95% para los coeficientes. Así, hay 95% de probabilidad de que el verdadero exponente de la pendiente esté entre 0.363 y 0.504, y de que el verdadero coeficiente del radio hidráulico esté entre 0.631 y 0.835. De esta forma, el ajuste no contradice los exponentes teóricos.

Finalmente, se debe observar que se puede usar la herramienta Solver de Excel para una *regresión no lineal*, minimizando de manera directa la suma de los cuadrados de los residuos entre una predicción del modelo no lineal y los datos. Dedicaremos la sección 20.1 a un ejemplo de cómo se realiza lo anterior.

19.8.2 MATLAB

Como se resume en la tabla 19.2, MATLAB tiene varias funciones preconstruidas que abarcan todas las capacidades que se describen en esta parte del libro. El siguiente ejemplo ilustra cómo usar algunas de ellas.

Tabla 19.2 Algunas funciones de MATLAB para implementar interpolación, regresión, trazadores y TRF.

Función	Descripción
`polyfit`	Ajusta polinomios a datos
`interp1`	Interpolación 1-D (tabla 1-D)
`interp2`	Interpolación 2-D (tabla 2-D)
`spline`	Interpolación de datos con splin cúbico
`fft`	Transformada discreta de Fourier

EJEMPLO 19.5 Uso de MATLAB para el ajuste de curvas

Planteamiento del problema Explore cómo se utiliza MATLAB para ajustar curvas a datos. Para ello, use la función seno para generar valores regularmente espaciados $f(x)$ de 0 a 10. Utilice un tamaño de paso de 1, de tal forma que la caracterización resultante de la onda seno sea dispersa (figura 19.22). Después, ajústela con interpolación *a*) lineal, *b*) polinomial de quinto grado y *c*) trazadora cúbica.

Solución

a) Los valores de las variables independientes y dependientes se introducen en vectores mediante

```
>> x=0:10;
>> y=sin(x);
```

Se genera un nuevo vector más finamente espaciado con valores de la variable independiente y se guarda en el vector **xi**,

```
>> xi=0:.25:10;
```

La función MATLAB `interp1` se usa después para generar valores de la variable dependiente *yi* para todos los valores *xi* usando interpolación lineal. Tanto los valores originales (x, y) como los valores interpolados linealmente se grafican juntos, como se muestra en la gráfica siguiente:

```
>> yi=interp1(x,y,xi);
>> plot(x,y,'o',xi,yi)
```

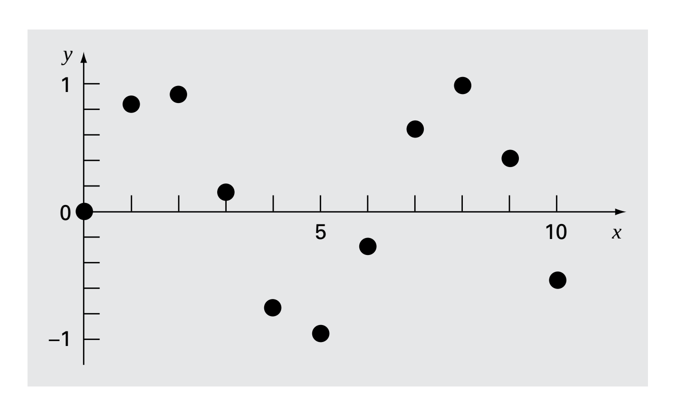
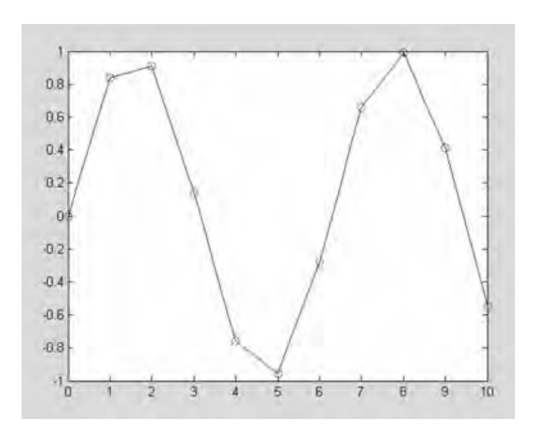


Figura 19.22
Once puntos muestreados de una sinusoide.

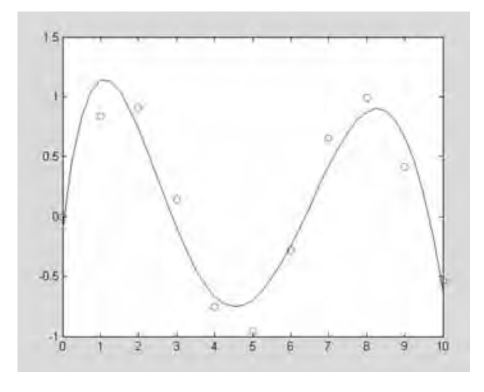


b) A continuación, la función `polyfit` de MATLAB se emplea para generar los coeficientes de un ajuste polinomial de quinto grado a los datos dispersos originales,

```
>> p=polyfit(x,y,5)
p=
  0.0008 -0.0290 0.3542 -1.6854 2.5860 -0.0915
```

donde el vector **p** contiene los coeficientes polinomiales. Éstos, a su vez, se utilizan para generar un nuevo conjunto de valores *yi*, los cuales de nuevo pueden graficarse junto con la muestra original dispersa,

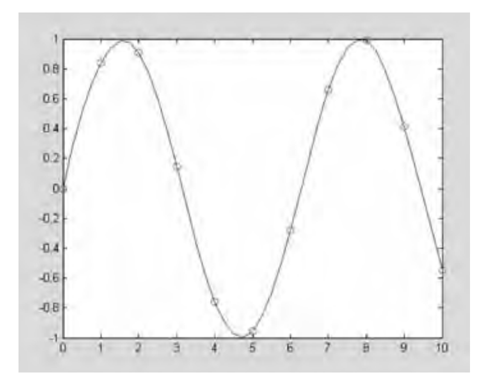
```
>> yi = polyval(p,xi);
>> plot(x,y,'o',xi,yi)
```



El polinomio captura el comportamiento general que siguen los datos; aunque deja fuera a la mayoría de los puntos.

c) Finalmente, la función `spline` de MATLAB puede servir para ajustar un trazador cúbico a los datos originales dispersos, en la forma de un nuevo conjunto de valores *yi*, los cuales nuevamente se grafican junto con la muestra original dispersa,

```
>> yi=spline(x,y,xi);
>> plot(x,y,'o',xi,yi)
```



MATLAB también tiene excelentes capacidades para realizar el análisis de Fourier. Se dedica la sección 20.3 a un ejemplo de cómo hacerlo.

MATLAB tiene dos funciones incorporadas para interpolación discreta bidimensional y tridimensional: `interp2` e `interp3`. Como podría esperarse por sus nombres, estas funciones operan de modo similar a `interp1`. Por ejemplo, una representación sencilla de la sintaxis de `interp2` es

```
zi = interp2(x, y, z, xi, yi, 'method')
```

donde `x` y `y` = matrices que contienen las coordenadas de los puntos en los que se dan los valores en la matriz `z`, `zi` = matriz que contiene los resultados de la interpolación evaluados en los puntos

en las matrices `xi` y `yi`, y `method` = el método que se desea. Observe que los métodos son idénticos a los que se usan por `interp1`; es decir, `linear`, `nearest`, `spline` y `cubic`.

Igual que con `interp1`, si se omite el argumento `method`, entra forzosamente interpolación lineal. Por ejemplo, se puede usar `interp2` para hacer la misma evaluación que en el ejemplo 18.11 como

```
>> x=[2 9];
>> y=[1 6];
>> z=[60 57.5;55 70];
>> interp2(x,y,z,5.25,4.8)

ans =
   61.2143
```

19.8.3 Mathcad

Mathcad puede realizar una amplia variedad de tareas estadísticas, de ajuste de curvas y de suavización de datos. Algunas de estas tareas son relativamente simples, como graficar histogramas y calcular resúmenes estadísticos de población como media, mediana, varianza, desviaciones estándar y coeficientes de correlación. Además, Mathcad contiene un número de funciones para realizar regresión. Las funciones **slope** e **intercept** devuelven como respuesta la pendiente y la intercepción de la línea de ajuste de regresión de mínimos cuadrados. La función **regress** se usa para regresión polinómica de grado *n* de un conjunto completo de datos. La función **loess** realiza regresión polinómica localizada sobre rangos parciales de los datos que usted puede especificar. La función **interp** se puede usar para obtener valores inmediatos de *y* de un ajuste de regresión para un punto *x* dado. Las funciones **regress** y **loess** pueden también realizar regresión polinómica multivariada. Mathcad también ofrece la función **linfit** que se emplea para modelar datos con una combinación lineal de funciones arbitrarias. Finalmente, la función **genfit** está disponible para casos en los que los coeficientes del modelo aparecen en forma arbitraria. En este caso, las ecuaciones no lineales más difíciles se deben resolver por iteración.

Mathcad también tiene capacidades considerables para interpolación. Puede predecir valores intermedios conectando puntos de datos conocidos ya sea con líneas rectas (interpolación lineal), usando la función **linterp**, o con interpolación por trazador cúbico usando la función **cspline**, **pspline** o **Ispline**. Estas funciones de trazador permiten que usted pruebe diferentes maneras de manejar la interpolación en los puntos extremos de los datos. La función **Ispline** genera una curva de spline que es una línea recta en los puntos extremos. La función **pspline** genera una curva de trazador que es una parábola en los puntos extremos. La función **cspline** genera una curva de trazador que es cúbica en los puntos finales. La función **interp** usa los resultados del ajuste de curva y entrega un valor interpolado de *y* dado un valor de *x*. Además, se puede realizar interpolación bidimensional cúbica de trazador mediante el paso de una superficie a través de una cuadrícula de puntos.

Realicemos un ejemplo que muestra cómo se emplea Mathcad para realizar una interpolación de spline (figura 19.23). Los datos que incluiremos sólo son algunos puntos espaciados uniformemente, muestreados a partir de una curva sinusoidal. Después de generar esta información, el símbolo de la definición y la función **lspline** se usan para calcular los coeficientes del trazador. Luego se desarrolla una función de interpolación, fija, con la función **interp**, con objeto de generar valores interpolados para valores específicos de *x*. Mathcad diseñó esta secuencia de operaciones de tal manera que no sea necesario recalcular los polinomios de interpolación cada vez que se desee una interpolación. Con las funciones en su sitio, usted puede interpolar en cualquier ubicación usando fit(x), como se muestra con $x = 2.5$. También puede usted construir una gráfica de los datos junto con el trazador interpolado como se muestra en la figura 19.23.

Como otro ejemplo para demostrar algunas de las capacidades de Mathcad para ajuste de curvas, se usará la función **fft** para análisis de Fourier como en la figura 19.24. La primera línea usa el símbolo de definición para crear *i* como un rango variable. En seguida se formula x_i usando la función Mathcad **rnd** para impartir una componente aleatoria a la señal sinusoidal. La gráfica de la señal se puede colocar en la hoja de cálculo haciendo clic en la ubicación deseada. Esto pone una retícula roja en esa ubicación. Luego use el menú desplegable Insert/Graph/X-Y Plot para colocar una gráfica

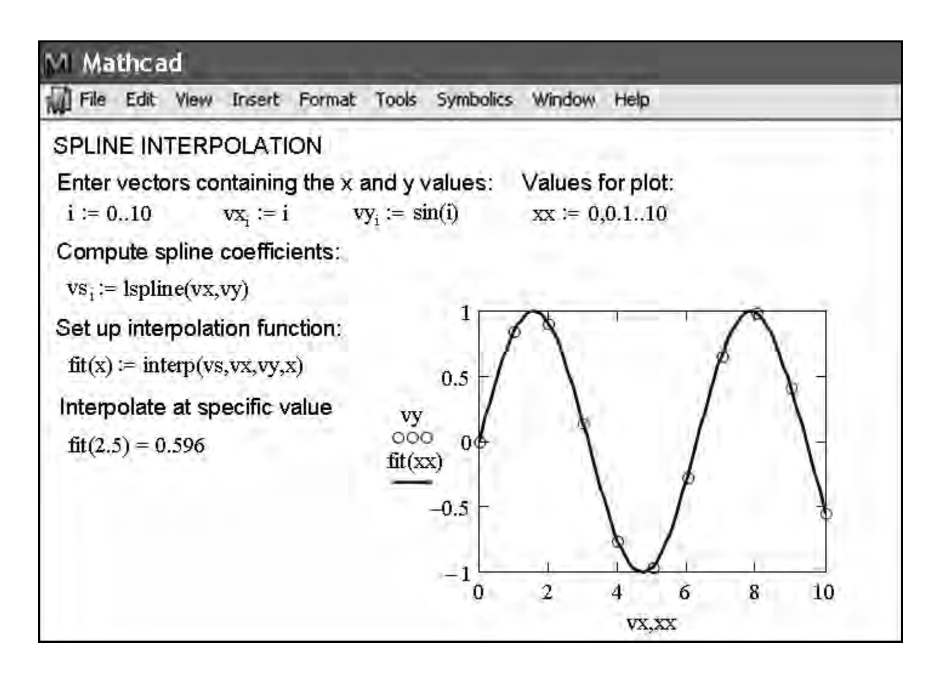


Figura 19.23
Interpolación de trazador cúbico con Mathcad.

vacía en la hoja de cálculo con espacios para introducir las expresiones que se van a graficar y para los rangos de los ejes *x* y *y*. Simplemente teclee x_i en el espacio indicado en el eje *y* e *i* para el eje *x*. Mathcad hace el resto para producir la primera gráfica que se muestra en la figura 19.24. Una vez que se ha creado la gráfica, usted puede usar el menú desplegable Format/Graph/X-Y Plot para variar el tipo de gráfica, cambiar el tipo de color, tipo y peso del trazo de la función, y agregar títulos, rótulos y otros atributos. Luego se define *c* como **fft**(x). Esta función entrega la transformada de Fourier de *x*. El resultado es un vector **c** de coeficientes complejos que representan valores en el dominio de la frecuencia. Entonces se construye una gráfica de la magnitud de c_j como se describe arriba.

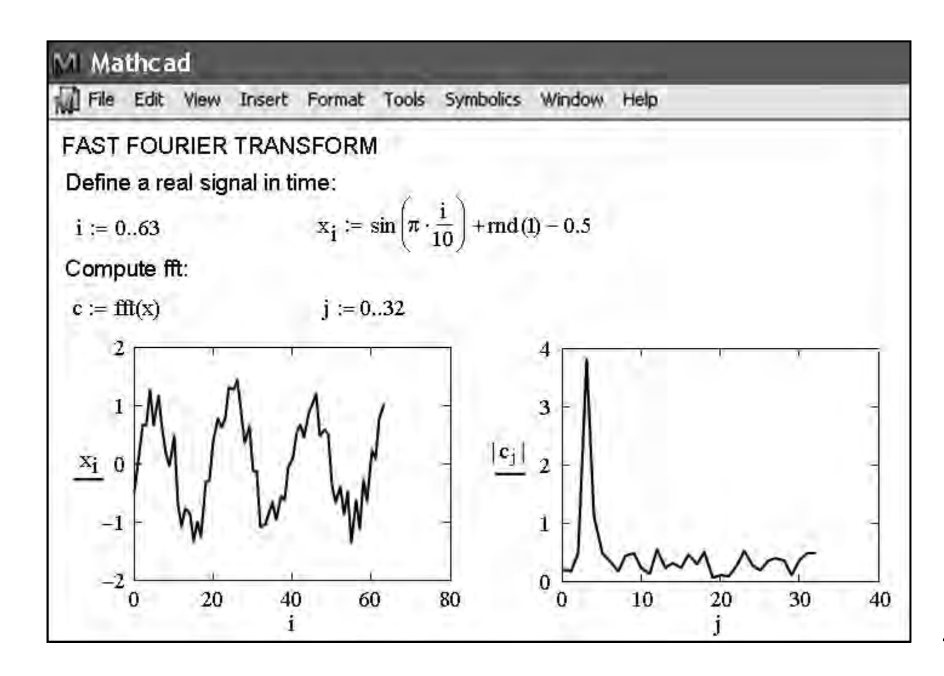


Figura 19.24
FFT (TRF) con Mathcad.

PROBLEMAS

19.1 Los valores promedio de una función se determinan por medio de

$$\overline{f(x)} = \frac{\int_0^x f(x)\,dx}{x}$$

Emplee esta relación para verificar los resultados de la ecuación (19.13).

19.2 Se ha tabulado la radiación solar en Tucson, Arizona, como sigue

Tiempo, meses	E	F	M	A	M	J
Radiación, W/m^2	144	188	245	311	351	359

Tiempo, meses	J	A	S	O	N	D
Radiación, W/m^2	308	287	260	211	159	131

Suponga que cada mes tiene 30 días y ajuste una sinusoide a estos datos. Utilice la ecuación resultante para pronosticar la radiación a mediados de agosto.

19.3 El pH en un reactor varía en formas sinusoidales durante el curso del día. Utilice regresión por mínimos cuadrados para ajustar la ecuación (19.11) a los datos siguientes. Use el ajuste para determinar la media, amplitud y tiempo del pH máximo. Note que el periodo es de 24 h.

Tiempo, h	0	2	4	5	7	9	12	15	20	22	24
pH	7.3	7	7.1	6.5	7.4	7.2	8.9	8.8	8.9	7.9	7

19.4 Use una serie de Fourier continua para aproximar la onda diente de sierra que se observa en la figura P19.4. Elabore la gráfica de los tres primeros términos junto con la suma.

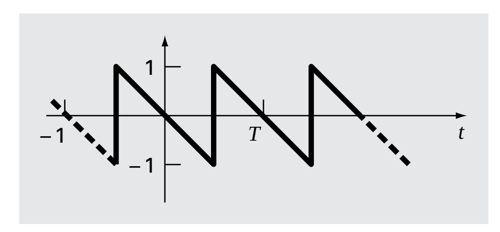


Figura P19.4
Onda diente de sierra.

19.5 Utilice una serie de Fourier continua para aproximar la forma de la onda que se ilustra en la figura P19.5. Grafique los tres primeros términos junto con la suma.

19.6 Construya los espectros de línea de amplitud y fase para el problema 19.4.

19.7 Construya los espectros de línea de amplitud y fase para el problema 19.5.

19.8 Un rectificador de media onda se caracteriza por medio de

$$C_1 = \left[\frac{1}{\pi} + \frac{1}{2}\,\text{sen}t - \frac{2}{3\pi}\cos 2t - \frac{2}{15\pi}\cos 4t - \frac{2}{35\pi}\cos 6t - \cdots\right]$$

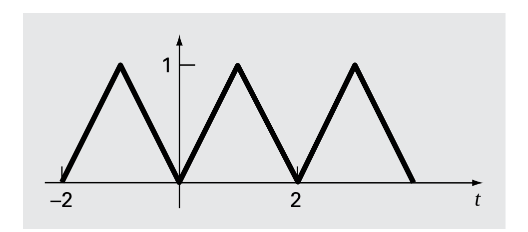


Figura P19.5
Onda triangular.

donde C_1 es la amplitud de onda. Grafique los primeros cuatro términos junto con la suma.

19.9 Construya espectros de línea de amplitud y fase para el problema 19.8.

19.10 Desarrolle un programa amigable para la TDF con base en el algoritmo de la figura 19.12. Pruébelo con la replicación de la figura 19.13.

19.11 Use el programa del problema 19.10 para calcular una TDF para la onda triangular del problema 19.8. Muestre la onda de $t = 0$ a $4T$. Use los puntos de muestra 32, 64 y 128. Tome el tiempo de cada corrida y haga la gráfica de la ejecución *versus* N para verificar la figura 19.14.

19.12 Desarrolle un programa amigable para la TRF con base en el algoritmo de la figura 19.18. Pruébelo con la duplicación de la figura 19.13.

19.13 Repita el problema 19.11 con el uso del software que desarrolló en el problema 19.12.

19.14 Un objeto está suspendido en un túnel de viento y se mide la fuerza para varios niveles de velocidad del viento. Los resultados se hallan tabulados a continuación. Use el comando de Excel Trendline para ajustar una ecuación de potencias a estos datos. Haga la gráfica de F *versus* v junto con la ecuación de potencias y r^2.

v, m/s	10	20	30	40	50	60	70	80
F, N	25	70	380	550	610	1 220	830	1 450

19.15 Use el paquete de herramientas para el análisis de datos de Excel para desarrollar un polinomio de regresión para los datos siguientes, para la concentración de oxígeno disuelto de agua dulce *versus* temperatura a nivel del mar. Determine el grado del polinomio necesario para alcanzar la precisión de los datos.

T, °C	0	8	16	24	32	40
o, mg/L	14.62	11.84	9.87	8.42	7.31	6.41

19.16 Use el paquete de herramienta para el análisis de datos de Excel para ajustar una línea recta a los datos siguientes. Determine el intervalo de confianza el 90% para la intersección. Si abarca al cero, vuelva a hacer la regresión, pero con la intersección forzada a ser cero (ésta es una opción en el cuadro de diálogo de **Regresión**).

x	2	4	6	8	10	12	14
y	6.5	7	13	17.8	19	25.8	26.9

19.17 *a*) Emplee MATLAB para ajustar un trazador cúbico a los datos siguientes:

x	0	2	4	7	10	12
y	20	20	12	7	6	6

Determine el valor de y en $x = 1.5$. *b*) Repita el inciso *a*), pero sin primeras derivadas en los nodos finales. Observe que la herramienta de ayuda de MATLAB describe cómo prescribir las derivadas finales.

19.18 Use MATLAB para generar 64 puntos de la función

$$f(t) = \cos(10t) + \text{sen}(3t)$$

de $t = 0$ a 2π. Con la función `randn` agregue un componente aleatorio a la señal. Tome una TRF de estos valores y grafique los resultados.

19.19 En forma similar a como se hizo en la sección 19.8.2, use MATLAB para ajustar los datos del problema 19.15 con *a*) interpolación lineal, *b*) un polinomio de regresión de tercer grado y *c*) un trazador. Use cada enfoque para predecir la concentración de oxígeno en $T = 10$.

19.20 La función de Runge es

$$f(x) = \frac{1}{1 + 25x^2}$$

Genere 9 valores equidistantes de esta función en el intervalo: [–1, 1]. Ajuste estos datos con *a*) un polinomio de grado ocho, *b*) un trazador lineal y *c*) un trazador cúbico. Presente sus resultados en forma gráfica.

19.21 Se inyecta un colorante al torrente circulatorio de un paciente para medir su salida cardiaca, que es la tasa de flujo volumétrico de la sangre del ventrículo izquierdo del corazón. En otras palabras, la salida cardiaca es el número de litros de sangre que el corazón bombea por minuto. Para una persona en reposo, la tasa puede ser de 5 o 6 litros por minuto. Si se trata de un maratonista durante la carrera, la salida cardiaca puede ser tan elevada como 30 L/min. Los datos siguientes muestran la respuesta de un individuo cuando se inyectan 5 mg de colorante en el sistema vascular.

Tiempo (s)	2	6	9	12	15	18	20	24
Concentración (mg/L)	0	1.5	3.2	4.1	3.4	2	1	0

Ajuste una curva polinomial a través de los puntos de los datos y use la función para aproximar la salida cardiaca del paciente, que se puede calcular con:

$$\text{Salida cardiaca} = \frac{\text{cantidad de colorante}}{\text{área bajo la curva}}\left(\frac{\text{L}}{\text{min}}\right)$$

19.22 En los circuitos eléctricos es común ver el comportamiento de la corriente en la forma de una onda cuadrada como se ilustra en la figura P19.22. Al resolver para la serie de Fourier a partir de

$$f(t) = \begin{cases} A_0 & 0 \le t \le T/2 \\ -A_0 & T/2 \le t \le T \end{cases}$$

se obtiene la serie de Fourier siguiente

$$f(t) = \sum_{n=1}^{\infty} \left(\frac{4A_0}{(2n-1)\pi}\right) \text{sen}\left(\frac{2\pi(2n-1)t}{T}\right)$$

Sea $A_0 = 1$ y $T = 0.25$ s. Grafique los seis primeros términos de la serie de Fourier individualmente, así como la suma de dichos seis términos. Si es posible, use un paquete como Excel o MATLAB.

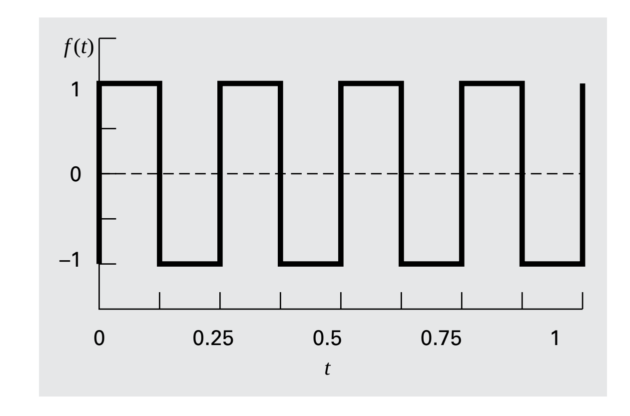


Figura P19.22

19.23 Haga una gráfica de los datos siguientes con *a*) un polinomio de interpolación de sexto grado, *b*) un trazador cúbico y *c*) un trazador cúbico con derivadas finales de cero.

x	0	100	200	400	600	800	1 000
f(x)	0	0.82436	1.00000	0.73576	0.40601	0.19915	0.09158

En cada caso, compare la gráfica con la ecuación siguiente, la cual se utilizó para generar los datos

$$f(x) = \frac{x}{200} e^{-\frac{x}{200}+1}$$

CAPÍTULO 20

Estudio de casos: ajuste de curvas

El propósito de este capítulo es usar los métodos numéricos para el ajuste de curvas en la solución de algunos problemas de ingeniería. La primera aplicación, tomada de la ingeniería química, muestra cómo un modelo no lineal se puede linealizar y ajustar a datos mediante regresión lineal. La segunda aplicación utiliza trazadores para estudiar un problema que tiene relevancia en el área ambiental de la ingeniería civil: transporte de calor y de masa en un lago estratificado.

La tercera aplicación ilustra cómo se emplea una transformada rápida de Fourier (TRF) en la ingeniería eléctrica, para analizar una señal determinando sus principales armónicas. La última aplicación muestra la forma en que se usa la regresión lineal múltiple para analizar datos experimentales en un problema de fluidos tomado de la ingeniería mecánica y aeronáutica.

20.1 REGRESIÓN LINEAL Y MODELOS DE POBLACIÓN (INGENIERÍA QUÍMICA/BIOINGENIERÍA)

Antecedentes Los modelos de crecimiento poblacional son importantes en diversos campos de la ingeniería. En muchos de los modelos es fundamental la hipótesis de que la razón de cambio de la población (dp/dt) es proporcional a la población existente (p) en cualquier tiempo (t), o en forma de ecuación,

$$\frac{dp}{dt} = kp \tag{20.1}$$

donde k = un factor de proporcionalidad conocido como velocidad de crecimiento específico y tiene las unidades de tiempo^{-1}. Si k es una constante, entonces la solución de la ecuación (20.1) se obtiene de la teoría de las ecuaciones diferenciales:

$$p(t) = p_0 e^{kt} \tag{20.2}$$

donde p_0 = población cuando $t = 0$. En la ecuación (20.2) se observa que $p(t)$ se aproxima al infinito conforme t crece. Tal comportamiento es claramente imposible en la realidad. Por lo tanto, el modelo debe modificarse para hacerlo más realista.

Solución Primero, se debe reconocer que la velocidad de crecimiento específico k no puede ser una constante conforme la población crece. Éste es el caso debido a que, conforme p se aproxima al infinito, el fenómeno que está modelando se verá limitado por algunos factores como por ejemplo carencia de alimentos y producción de desechos tóxicos. Una manera de expresar esto en forma matemática es mediante el uso de un modelo de velocidad de crecimiento de saturación tal que

$$k = k_{\text{máx}} = \frac{f}{K + f} \tag{20.3}$$

donde $k_{\text{máx}}$ = la *velocidad de crecimiento máximo obtenible* para valores grandes de alimento (f) y K = la *constante de saturación media*. En la figura 20.1 se tiene la gráfica de la ecuación (20.3) y se muestra que cuando $f = K$, $k = k_{\text{máx}}/2$. Por lo tanto, K es la cantidad de alimento disponible que permite una velocidad de crecimiento poblacional igual a la mitad de la velocidad máxima.

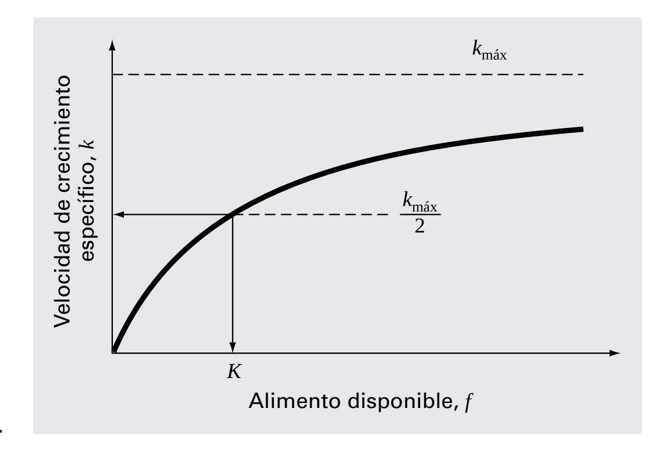


Figura 20.1
Gráfica de la velocidad de crecimiento específico contra alimento disponible para el modelo de velocidad de crecimiento de saturación usado para caracterizar la cinética microbiana. El valor K se conoce como constante de saturación media porque concuerda con la concentración a la que la velocidad de crecimiento específico es la mitad de su valor máximo.

Las constantes K y $k_{máx}$ son valores empíricos obtenidos de mediciones experimentales de k para diversos valores de f. Por ejemplo, suponga que la población p representa la levadura empleada en la producción comercial de cerveza, y f es la concentración de la fuente de carbono que será fermentada. Las mediciones de k contra f para la levadura se muestran en la tabla 20.1. Se requiere calcular $k_{máx}$ y K a partir de estos datos empíricos. Esto se logra invirtiendo la ecuación (20.3) de manera similar a la ecuación (17.17) para obtener

$$\frac{1}{k} = \frac{K+f}{k_{máx}\, f} = \frac{K}{k_{máx}}\frac{1}{f} + \frac{1}{k_{máx}} \qquad (20.4)$$

Con esta manipulación se ha transformado la ecuación (20.3) a una forma lineal; es decir, $1/k$ es una función lineal de $1/f$, con pendiente $K/k_{máx}$ e intersección $1/k_{máx}$. Estos valores se grafican en la figura 20.2.

A causa de dicha transformación, se pueden utilizar los métodos por mínimos cuadrados lineales, descritos en el capítulo 17, para determinar $k_{máx}$ = 1.23 días^{-1} y K = 22.18 mg/L. Los resultados, combinados con la ecuación (20.3), se comparan con los datos no transformados en la figura 20.3, y cuando se sustituyen en el modelo de la ecuación (20.1) dan el resultado siguiente:

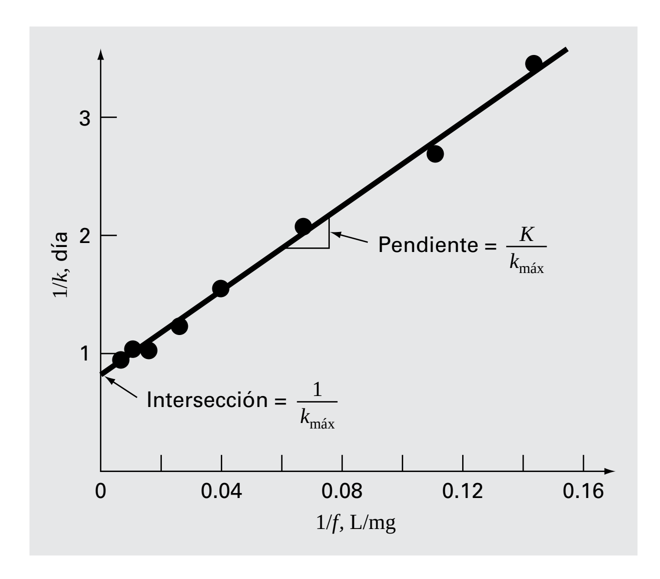


Figura 20.2
Versión linealizada del modelo de la velocidad de crecimiento de saturación. La línea es un ajuste por mínimos cuadrados que se utiliza para evaluar los coeficientes del modelo $k_{máx}$ = 1.23 días^{-1} y k = 22.18 mg/L para una levadura que sirve para producir cerveza.

Tabla 20.1 Datos utilizados para evaluar las constantes de un modelo de velocidad de crecimiento de saturación para caracterizar la cinética microbiana.

f, mg/L	k, día^{-1}	$1/f$, L/mg	$1/k$, día
7	0.29	0.14286	3.448
9	0.37	0.11111	2.703
15	0.48	0.06666	2.083
25	0.65	0.04000	1.538
40	0.80	0.02500	1.250
75	0.97	0.01333	1.031
100	0.99	0.01000	1.010
150	1.07	0.00666	0.935

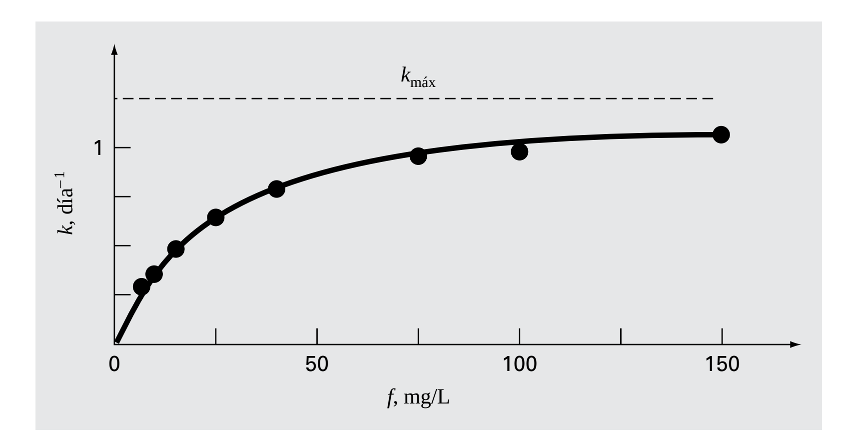


Figura 20.3
Ajuste del modelo de velocidad de crecimiento de saturación para una levadura empleada en la producción comercial de cerveza.

$$\frac{dp}{dt} = 1.23 \frac{f}{22.18 + f} p \tag{20.5}$$

Observe que el ajuste da una suma de los cuadrados de los residuos (como se calculó con los datos no transformados) es de 0.001305.

La ecuación (20.5) se resuelve usando la teoría de las ecuaciones diferenciales o los métodos numéricos que se analizan en el capítulo 25 cuando se conoce $f(t)$. Si f se aproxima a cero conforme p crece, entonces dp/dt se aproxima a cero y se estabiliza la población.

La linealización de la ecuación (20.3) constituye una forma para evaluar las constantes $k_{máx}$ y K. Un procedimiento alternativo, que se ajusta a la relación en su forma original, es la regresión no lineal descrita en la sección 17.5. La figura 20.4 muestra cómo se emplea la herramienta Solver de Excel para estimar los parámetros con la regresión no lineal. Como se observa, se desarrolla una columna de valores predichos basada en el modelo y en los parámetros iniciales. Éstos se utilizan para generar una columna de residuos al cuadrado que se suman, y el resultado se coloca en la celda D14. Después se usa el Solver de Excel para minimizar la celda D14 al ajustar las celdas B1:B2. El resultado, como se muestra en la figura 20.4, da estimados de $k_{máx} = 1.23$ y $K = 22.14$, con un $S_r = 0.001302$. De esta forma, aunque, como se esperaba, la regresión no lineal ofrece un ajuste ligeramente mejor, los resultados son casi idénticos. En otros casos, esto puede no ser así (o la función quizá no sea compatible con linealización) y la regresión no lineal podría ser la única opción factible para obtener un ajuste por mínimos cuadrados.

	A	B	C	D
1	kmáx	1.2301		
2	K	22.1386		
3				
4	f	k	k-predicción	Res^2
5	7	0.29	**0.295508**	**0.000030**
6	9	0.37	0.355536	0.000209
7	15	0.48	0.496828	0.000283
8	25	0.65	0.652385	0.000006
9	40	0.8	0.791843	0.000067
10	75	0.97	0.949751	0.000410
11	100	0.99	1.007135	0.000294
12	150	1.07	1.071898	0.000004
13				
14			SSR	**0.001302**

=$B$1*A5/($B$2+A5)

=(B5-C5)^2

=SUM(D5..D12)

Figura 20.4
Regresión no lineal para ajustar el modelo de la velocidad de crecimiento de saturación de una levadura empleada en la producción comercial de cerveza.

20.2 USO DE TRAZADORES PARA ESTIMAR LA TRANSFERENCIA DE CALOR (INGENIERÍA CIVIL/AMBIENTAL)

Antecedentes Los lagos de la zona templada llegan a dividirse en estratos térmicos durante el verano. Como se ilustra en la figura 20.5, cerca de la superficie, el agua es tibia y ligera, y en el fondo es más fría y densa. La estratificación divide efectivamente el lago en dos capas en forma vertical: el *epilimnion* y el *hipolimnion*, separadas por un plano conocido como *termoclina*.

La estratificación térmica tiene gran importancia para los ingenieros ambientales que estudian la contaminación de tales sistemas. En particular, la *termoclina* disminuye en gran medida la mezcla de las dos capas. Como resultado, la descomposición de la materia orgánica puede conducir a una gran reducción de oxígeno en el fondo aislado de las aguas.

La ubicación de la *termoclina* se puede definir como el punto de inflexión de la curva temperatura-profundidad; es decir, el punto donde $d^2T/dx^2 = 0$. Es también el punto en el cual el valor absoluto de la primera derivada o gradiente es un máximo. Utilice trazadores cúbicos para determinar la profundidad de la *termoclina* en el lago Platte (tabla 20.2). También use los trazadores para determinar el valor del gradiente en la *termoclina*.

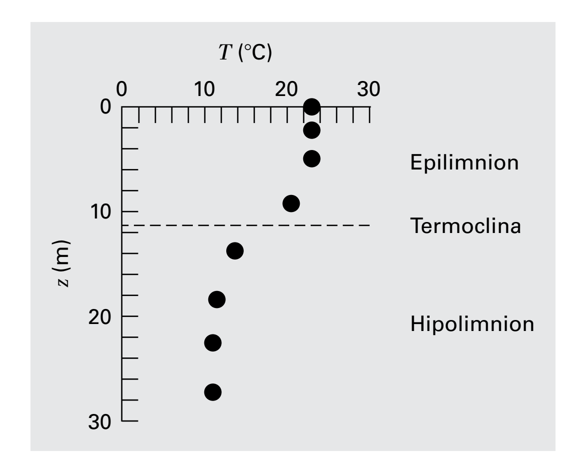


Figura 20.5
Temperatura contra profundidad durante el verano en el lago Platte, Michigan.

Solución Los datos se analizan con un programa que se desarrolló con base en el pseudocódigo de la figura 18.18. Los resultados se muestran en la tabla 20.3 que da las predicciones del trazador junto con las primera y segunda derivadas a intervalos de 1 m hacia abajo a través de la columna de agua.

Los resultados se grafican en la figura 20.6. Observe cómo la *termoclina* está claramente localizada en la profundidad donde el gradiente es mayor (es decir, el valor absoluto de la derivada es mayor) y la segunda derivada es cero. La profundidad es 11.35 m y el gradiente en este punto es –1.61°C/m.

Tabla 20.2 Temperatura contra profundidad durante el verano en el lago Platte, Michigan.

T, °C	22.8	22.8	22.8	20.6	13.9	11.7	11.1	11.1
z, m	0	2.3	4.9	9.1	13.7	18.3	22.9	27.2

Tabla 20.3 Resultados del programa trazadores basado en el pseudocódigo de la figura 18.18.

Profundidad (m)	T(C)	dT/dz	d2T/dz2	Profundidad (m)	T(C)	dT/dz	d2T/dz2
0.	22.8000	–.0115	.0000	15.	12.7652	–.6518	.3004
1.	22.7907	–.0050	.0130	16.	12.2483	–.3973	.2086
2.	22.7944	.0146	.0261	17.	11.9400	–.2346	.1167
3.	22.8203	.0305	–.0085	18.	11.7484	–.1638	.0248
4.	22.8374	–.0055	–.0635	19.	11.5876	–.1599	.0045
5.	22.7909	–.0966	–.1199	20.	11.4316	–.1502	.0148
6.	22.6229	–.2508	–.1884	21.	11.2905	–.1303	.0251
7.	22.2665	–.4735	–.2569	22.	11.1745	–.1001	.0354
8.	21.6531	–.7646	–.3254	23.	11.0938	–.0596	.0436
9.	20.7144	–1.1242	–.3939	24.	11.0543	–.0212	.0332
10.	19.4118	–1.4524	–.2402	25.	11.0480	.0069	.0229
11.	17.8691	–1.6034	–.0618	26.	11.0646	.0245	.0125
12.	16.2646	–1.5759	.1166	27.	11.0936	.0318	.0021
13.	14.7766	–1.3702	.2950	28.	11.1000	.0000	.0000
14.	13.5825	–.9981	.3923				

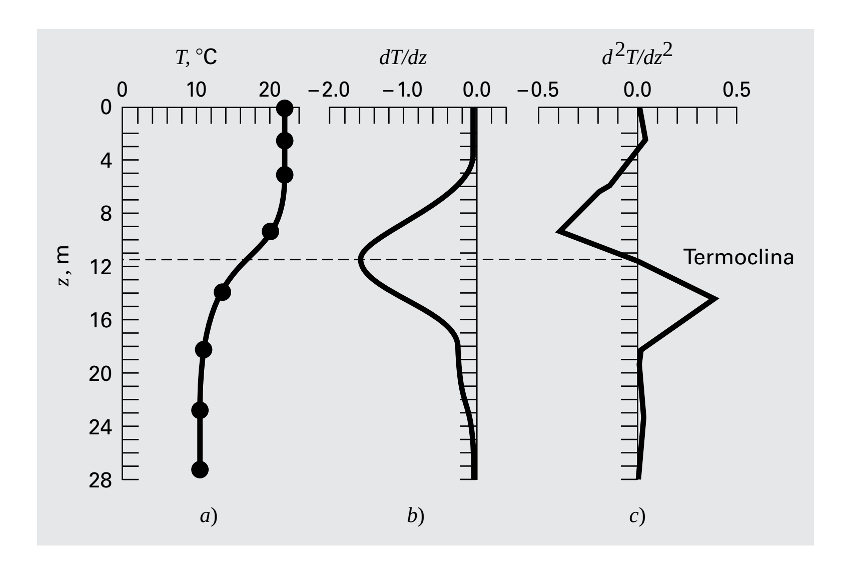


Figura 20.6
Gráficas de *a*) temperatura, *b*) gradiente y *c*) segunda derivada contra profundidad (*m*) generadas con el programa de trazadores cúbicos. La termoclina se localiza en el punto de inflexión de la curva temperatura-profundidad.

20.3 ANÁLISIS DE FOURIER (INGENIERÍA ELÉCTRICA)

Antecedentes El análisis de Fourier se emplea en muchas áreas de la ingeniería. Se utiliza de manera extensiva en problemas de la ingeniería eléctrica como el procesamiento de señales.

En 1848, Johann Rudolph Wolf diseñó un método para cuantificar la actividad solar contando el número de manchas y grupos de manchas en la superficie solar. Calculó una cantidad, que ahora se conoce como el *número de manchas solares de Wolf*, sumando 10 veces el número de grupos más el número de manchas solares. Como se observa en la figura 20.7, el registro de este número se remonta a 1700. Basándose en los primeros datos históricos, Wolf determinó que la longitud del ciclo es de 11.1 años.

Use un análisis de Fourier para confirmar este resultado mediante la aplicación de una TRF a los datos de la figura 20.7. Determine con toda precisión el periodo desarrollando una gráfica de potencia contra periodo.

Solución Los datos de años y el número de manchas solares se bajaron de Internet[1] y se guardaron en un archivo llamado: sunspot.dat. El archivo se puede cargar en MATLAB y la información del año y el número se le asignó a vectores con los mismos nombres,

```
>> load sunspot.dat
>> year=sunspot(:,1);number=sunspot(:,2);
```

A continuación, se aplica una TRF a los números de manchas solares

```
>> y=fft(number);
```

Una vez obtenida la primera armónica, se determina la longitud de la transformada rápida de Fourier (n) y luego se calculan la potencia y la frecuencia,

```
>> y(1)=[ ];
>> n=length(y);
>> power=abs(y(1:n/2)).^2;
>> nyquist=1/2;
>> freq=(1:n/2)/(n/2)*nyquist;
```

[1] Al momento de la impresión de la edición en inglés de este libro la página era http://www.ngdc.noaa.gov//stp/SOLAR/SSN/ssn.html.

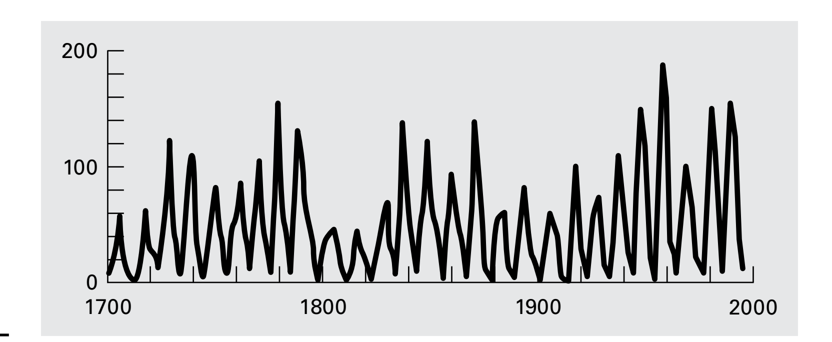


Figura 20.7
Gráfica del número de manchas solares de Wolf contra años.

En este momento, el espectro de potencia es una gráfica de potencia contra frecuencia. Sin embargo, como el periodo es más significativo en el contexto presente, se puede determinar el periodo y una gráfica potencia-periodo,

```
>> period=1./freq
>> plot(period,power);
```

El resultado, como se muestra en la figura 20.8, indica un pico alrededor de 11 años. El valor exacto se calcula con

```
>> index=find(power==max(power));
>> period(index)

ans=
  10.9259
```

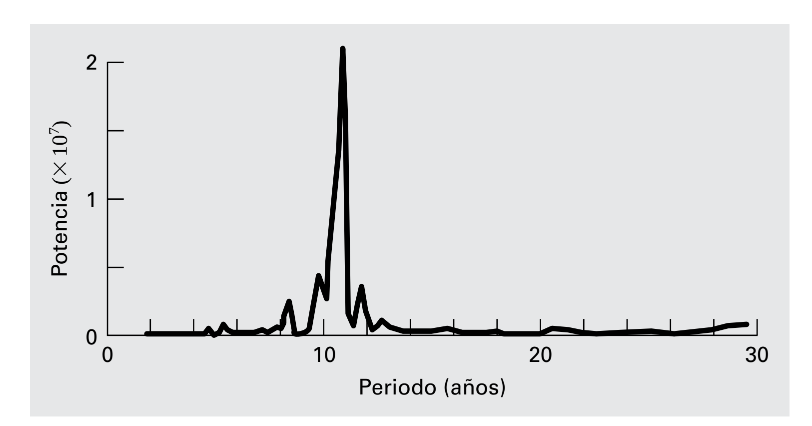


Figura 20.8
Espectro de potencia para el número de manchas solares de Wolf.

20.4 ANÁLISIS DE DATOS EXPERIMENTALES (INGENIERÍA MECÁNICA/AERONÁUTICA)

Antecedentes Las variables de diseño en la ingeniería son a menudo dependientes de varias variables independientes. Por lo común, esta dependencia funcional se caracteriza mejor con ecuaciones de potencia multivariable. Como se analizó en la sección 17.3, una regresión lineal múltiple de datos transformados a logaritmos ofrece un recurso para evaluar tales relaciones.

Por ejemplo, un estudio en ingeniería mecánica indica que el flujo de un líquido a través de una tubería está relacionado con el diámetro y la pendiente de la tubería (tabla 20.4). Use regresión lineal múltiple para analizar estos datos. Después con el modelo resultante prediga el flujo en una tubería de 2.5 ft de diámetro y con pendiente de 0.025 ft/ft.

Solución La ecuación de potencias a evaluarse es

$$Q = a_0 D^{a_1} S^{a_2} \tag{20.6}$$

donde Q = flujo (ft^3/s), S = pendiente (ft/ft), D = diámetro de la tubería (ft) y a_0, a_1 y a_2 = coeficientes. Tomando los logaritmos de esta ecuación se obtiene

$$\log Q = \log a_0 + a_1 \log D + a_2 \log S$$

En esta forma, la ecuación es adecuada para una regresión lineal múltiple, ya que log Q es una función lineal de log S y log D. Usando el logaritmo (base 10) de los datos de la tabla 20.4, se generan las siguientes ecuaciones expresadas en forma matricial [ecuación (17.22)]:

Tabla 20.4 Datos experimentales de diámetro, pendiente y flujo en una tubería circular de concreto.

Experimento	Diámetro, ft	Pendiente, ft/ft	Flujo, ft³/s
1	1	0.001	1.4
2	2	0.001	8.3
3	3	0.001	24.2
4	1	0.01	4.7
5	2	0.01	28.9
6	3	0.01	84.0
7	1	0.05	11.1
8	2	0.05	69.0
9	3	0.05	200.0

$$\begin{bmatrix} 9 & 2.334 & -18.903 \\ 2.334 & 0.954 & -4.903 \\ -18.903 & -4.903 & 44.079 \end{bmatrix} \begin{Bmatrix} \log a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} 11.691 \\ 3.945 \\ -22.207 \end{Bmatrix}$$

Este sistema se resuelve utilizando la eliminación de Gauss para obtener

$$\log a_0 = 1.7475$$
$$a_1 = 2.62$$
$$a_2 = 0.54$$

Si $\log a_0 = 1.7475$, entonces $a_0 = 10^{1.7475} = 55.9$, y la ecuación (20.6) ahora es

$$Q = 55.9D^{2.62}S^{0.54} \tag{20.7}$$

La ecuación (20.7) se utiliza para predecir el flujo para el caso de $D = 2.5$ ft y $S = 0.025$ ft/ft, como sigue

$$Q = 55.9(2.5)^{2.62}(0.025)S^{0.54} = 84.1 \text{ ft}^3/\text{s}$$

Debe observarse que la ecuación (20.7) se utiliza para otros propósitos, además del cálculo de flujo. Por ejemplo, la pendiente se relaciona con la pérdida de presión h_L y la longitud de tubería L mediante $S = h_L/L$. Si esta relación se sustituye en la ecuación (20.7) y en la fórmula resultante se despeja h_L, se obtiene la siguiente ecuación:

$$h_L = \frac{L}{1\,721} Q^{1.85} D^{4.85}$$

Esta relación se conoce como *ecuación de Hazen-Williams*.

PROBLEMAS

Ingeniería química/bioingeniería

20.1 Desarrolle el mismo cálculo que en la sección 20.1, pero use regresión lineal y transformaciones para ajustar los datos con una ecuación de potencias. Evalúe el resultado.

20.2 Usted lleva a cabo experimentos y determina los valores siguientes de capacidad calorífica c a distintas temperaturas T para un gas:

T	–50	–30	0	60	90	110
c	1 270	1 280	1 350	1 480	1 580	1 700

Use regresión para determinar un modelo para predecir c como función de T.

20.3 Se sabe que el esfuerzo a la tensión de un plástico se incrementa como función del tiempo que recibe tratamiento a base de calor. Se obtuvieron los datos siguientes:

Tiempo	10	15	20	25	40	50	55	60	75
Esfuerzo a la tensión	5	20	18	40	33	54	70	60	78

a) Ajuste una línea recta a estos datos y utilice la ecuación para determinar el esfuerzo a la tensión en un tiempo de 32 min.

b) Repita el análisis para una línea recta con intersección en el origen.

20.4 Los datos siguientes se recabaron para determinar la relación entre la presión y la temperatura de un volumen fijo de 1 kg de nitrógeno. El volumen es de 10 m^3.

T, °C	–40	0	40	80	120	160
p, N/m^2	6 900	8 100	9 300	10 500	11 700	12 900

Emplee la ley del gas ideal $pV = nRT$ para determinar R sobre la base de dichos datos. Observe que para la ley, T debe expresarse en kelvins.

20.5 El volumen específico de un vapor sobrecalentado se enlista en tablas de vapor para distintas temperaturas. Por ejemplo, a una presión absoluta de 3 000 lb/in^2:

T, °F	700	720	740	760	780
v, ft^3/lb$_m$	0.0977	0.12184	0.14060	0.15509	0.16643

Determine v con T = 750°F.

20.6 Un reactor está estratificado térmicamente en la tabla siguiente:

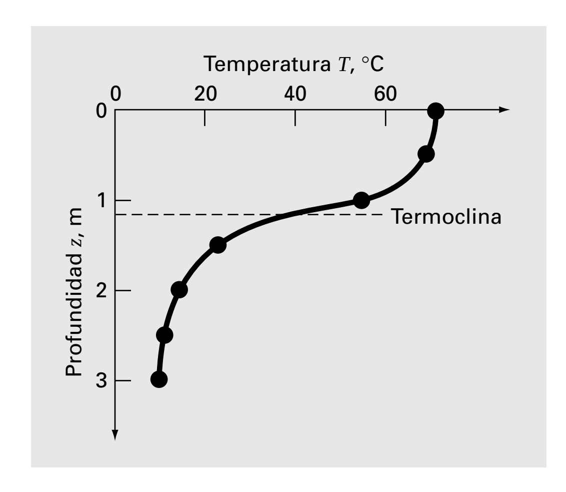


Figura P20.6

Profundidad, m	0	0.5	1.0	1.5	2.0	2.5	3.0
Temperatura, °C	70	68	55	22	13	11	10

Como se ilustra en la figura P20.6, el tanque puede idealizarse como dos zonas separadas por un gradiente fuerte de temperatura, o termoclina. La profundidad de este gradiente se define como el punto de inflexión de la curva temperatura-profundidad, es decir, el punto en el que $d^2T/dz^2 = 0$. A esta profundidad, el flujo de calor de la superficie a la capa del fondo se calcula con la ley de Fourier:

$$J = -k\frac{dT}{dz}$$

Use un ajuste con trazadores cúbicos de estos datos para determinar la profundidad de la termoclina. Si k = 0.02 cal/(s · cm · °C), calcule el flujo a través de esta interfaz.

20.7 En la enfermedad de Alzheimer, el número de neuronas en la corteza disminuye conforme la enfermedad avanza. Los datos siguientes se tomaron para determinar el número de receptores neurotransmisores que quedan en un cerebro enfermo. Se incubaron neurotransmisores libres ([F]) con tejido, y se midió la concentración que limita específicamente a un receptor ([B]). Cuando la cubierta es específica de un receptor, la concentración límite se relaciona con la concentración libre por medio de la relación siguiente:

$$[B] = \frac{B_{\text{máx}}[F]}{K + [F]}$$

Con el uso de los datos siguientes, determine los parámetros que minimizan la suma de los cuadrados de los residuos. Asimismo, calcule r^2.

[F], nM	0.1	0.5	1	5	10	20	50
[B], nM	10.57	36.61	52.93	82.65	89.46	94.35	101.00

20.8 Se tomaron los datos siguientes del tanque de un reactor de agitación para la reacción A → B. Use los datos para hacer las estimaciones mejores posibles para k_{01} y E_1, para el modelo cinético siguiente,

$$-\frac{dA}{dt} = k_{01}e^{-\frac{E_1}{RT}}A$$

donde R es la constante de los gases y es igual a 0.00198 Kcal/mol/K

$-dA/dt$ (moles/L/s)	460	960	2 485	1 600	1 245
A (moles/L)	200	150	50	20	10
T (K)	280	320	450	500	550

20.9 Emplee el conjunto siguiente de datos de presión-volumen para encontrar las mejores constantes viriales posibles (A_1 y A_2) para la ecuación de estado que se muestra a continuación. R = 82.05 ml atm/gmol K y T = 303 K.

$$\frac{PV}{RT} = 1 + \frac{A_1}{V} + \frac{A_2}{V^2}$$

P (atm)	0.985	1.108	1.363	1.631
V (ml)	25 000	22 200	18 000	15 000

20.10 Se tomaron datos de concentración en 15 puntos temporales para la reacción de polimerización:

$$xA + yB \rightarrow A_xB_y$$

Se supone que la reacción ocurre a través de un mecanismo complejo que consiste en muchas etapas. Se han planteado varios modelos hipotéticos y calculado la suma de los cuadrados de los residuos para los ajustes de los modelos a los datos. A continuación se presentan los resultados. ¿Cuál es el modelo que describe mejor los datos (estadísticamente)? Explique su respuesta.

	Modelo A	**Modelo B**	**Modelo C**
S_r	135	105	100
Número de modelo parámetros del ajuste	2	3	5

20.11 A continuación se presentan datos de la vasija de un reactor de crecimiento bacterial (una vez que terminó la fase de retraso). Se permite que las bacterias crezcan tan rápido como sea posible durante las primeras 2.5 horas, y después se les induce a producir una

proteína recombinante, la cual disminuye el crecimiento bacterial en forma significativa. El crecimiento teórico de las bacterias se describe por medio de:

$$\frac{dX}{dt} = \mu X$$

donde X es el número de bacterias, y μ es la tasa de crecimiento específico de las bacterias durante el crecimiento exponencial. Con base en los datos, estime la tasa de crecimiento específico de las bacterias durante las primeras 2 horas de crecimiento, así como durante las siguientes 4 horas de crecimiento.

Tiempo, h	0	1	2	3	4	5	6
[Células], g/L	0.100	0.332	1.102	1.644	2.453	3.660	5.460

20.12 El peso molecular de un polímero se determina a partir de su viscosidad por medio de la relación siguiente:

$$[\eta] = KM_v^a$$

donde $[\eta]$ es la viscosidad intrínseca del polímero, M_v es la viscosidad promediada del peso molecular, y K y a son constantes específicas del polímero. La viscosidad intrínseca se determina en forma experimental por medio de determinar el tiempo de flujo, o el tiempo que toma a la solución polimérica fluir entre dos líneas grabadas en un viscosímetro capilar, a distintas concentraciones de polímero diluido, y se extrapola para una dilución infinita. La gráfica de

$$\frac{\frac{t}{t_0} - 1}{c} \text{ versus } c$$

debe generar una línea recta, con intersección en el eje y igual a $[\eta]$. La concentración de la solución polimérica es c, t es el tiempo de flujo de la solución polimérica, y t_0 es el tiempo de flujo del solvente sin polímero. Con el uso de los datos siguientes de tiempos de flujo, para soluciones diluidas de poliestireno en metil etil acetona a 25°C, y las constantes $K = 3.9 \times 10^{-4}$, y $a = 0.58$, encuentre el peso molecular de la muestra de poliestireno.

Concentración de polímero, g/dL	Tiempo de flujo, s
0 (solvente puro)	83
0.04	89
0.06	95
0.08	104
0.10	114
0.12	126
0.14	139
0.16	155
0.20	191

20.13 En promedio, el área superficial A de los seres humanos se relaciona con el peso W y la estatura H. En la tabla siguiente se presentan los valores de A que se obtuvieron con mediciones de cierto número de individuos:

H (cm)	182	180	179	187	189	194	195	193	200
W (kg)	74	88	94	78	84	98	76	86	96
A (m²)	1.92	2.11	2.15	2.02	2.09	2.31	2.02	2.16	2.31

Desarrolle una ecuación para pronosticar el área como función de la estatura y el peso. Utilícela para estimar el área superficial de una persona de 187 cm y 78 kg.

20.14 Determine una ecuación para predecir la tasa del metabolismo como función de la masa con base en los datos siguientes:

Animal	Masa, kg	Metabolismo, watts
Vaca	400	270
Humano	70	82
Oveja	45	50
Gallina	2	4.8
Rata	0.3	1.45
Paloma	0.16	0.97

20.15 La sangre humana se comporta como un fluido newtoniano (véase el problema 20.55) en la región de la tasa de corte alto, donde $\dot{\gamma} > 100$. En la región de tasa de corte bajo, donde $\dot{\gamma} < 50$, los glóbulos rojos tienden a agregarse en lo que se denomina rouleaux (rodillos), que hacen que el comportamiento del fluido ya no sea newtoniano. Esta región de corte bajo se denomina región de Casson, y es una región de transición entre las dos regiones de flujo distinto. En la región de Casson, conforme la tasa de corte se aproxima a cero, el esfuerzo cortante adquiere un valor finito, similar al plástico Bingham, lo que se denomina esfuerzo inducido, τ_y, el cual debe superarse a fin de iniciar el flujo en la sangre estancada. El flujo en la región de Casson por lo general se grafica como la raíz cuadrada de la tasa de corte *versus* la raíz cuadrada del esfuerzo cortante, y sigue una relación lineal al graficarse de este modo. La relación de Casson es

$$\sqrt{\tau} = \sqrt{\tau_y} + K_c\sqrt{\dot{\gamma}}$$

donde K_c = índice de consistencia. En la tabla siguiente se muestran valores medidos en forma experimental de $\dot{\gamma}$ y τ, para una sola muestra de sangre en las regiones de Casson y de flujo newtoniano.

$\dot{\gamma}$, 1/s	0.91	3.3	4.1	6.3	9.6	23	36
τ, N/m²	0.059	0.15	0.19	0.27	0.39	0.87	1.33
Región	Casson						

$\dot{\gamma}$, 1/s	49	65	105	126	215	315	402
τ, N/m²	1.65	2.11	3.44	4.12	7.02	10.21	13.01
Región	Transición		Newtoniano				

Encuentre los valores de K_c y τ_y por medio de regresión lineal en la región de Casson, y halle μ con regresión lineal en la región newtoniana. También calcule el coeficiente de correlación para cada análisis de regresión. Grafique las dos rectas de regresión en una gráfica de Casson ($\sqrt{\dot{\gamma}}$ *versus* $\sqrt{\tau}$) y extienda las rectas de regresión como líneas punteadas hacia las regiones adyacentes; también incluya los puntos de los datos en la gráfica. Limite la región de la tasa de corte a $0 < \sqrt{\dot{\gamma}} < 15$.

20.16 El tejido suave sigue un comportamiento exponencial ante la deformación por tensión uniaxial, mientras esté en el rango fisiológico o normal de elongación. Esto se expresa como

$$\sigma = \frac{E_0}{a}(e^{a\varepsilon} - 1)$$

donde σ = esfuerzo, ε = tensión, y E_0 y a son constantes del material que se determinan en forma experimental. Para evaluar las dos constantes del material, la ecuación anterior se diferencia con respecto a ε. El uso de la ecuación establece la relación fundamental para el tejido suave

$$\frac{d\sigma}{d\varepsilon} = E_0 + a\sigma$$

Para evaluar E_0 y a, se grafican los datos de esfuerzo-tensión como $d\sigma/d\varepsilon$ *versus* σ, y la intersección y la pendiente de esta gráfica son las dos constantes del material, respectivamente.

En la tabla siguiente se muestran datos de esfuerzo-tensión para los tendones cordados del corazón (tendones pequeños que se usan para mantener cerradas las válvulas del corazón durante la contracción del músculo cardiaco; estos datos son para tejido que se carga, mientras que la descarga produce curvas diferentes).

σ, 10^3 N/m^2	87.8	96.6	176	263	351	571	834
ε, 10^{-3} m/m	153	204	255	306	357	408	459

σ, 10^3 N/m^2	1 229	1 624	2 107	2 678	3 380	4 258
ε, 10^{-3} m/m	510	561	612	663	714	765

Calcule la derivada $d\sigma/d\varepsilon$ con el uso de diferencias finitas. Grafique los datos y elimine los puntos de los datos cerca de los ceros que parezcan no seguir la relación de línea recta. El error en dichos datos proviene de la incapacidad de los instrumentos para leer los valores pequeños en esta región. Ejecute un análisis de regresión de los datos restantes a fin de determinar los valores de E_o y a.

Grafique los puntos del esfuerzo *versus* los de tensión junto con la curva analítica expresada por la primera ecuación. Esto indicará qué tan bien la curva analítica concuerda con los datos.

Muchas veces esto no funciona bien debido a que el valor de E_o es difícil de evaluar con esta técnica. Para resolver este problema, no se utiliza E_o. Se selecciona un punto de los datos $(\bar{\sigma}, \bar{\varepsilon})$ a la mitad del rango del análisis de regresión. Dichos valores se sustituyen en la primera ecuación y se determina un valor de E_o/a, el cual se sustituye en la primera ecuación, que se convierte en

$$\sigma = \left(\frac{\bar{\sigma}}{e^{a\bar{\varepsilon}} - 1}\right)(e^{a\varepsilon} - 1)$$

Con este enfoque, los datos experimentales que están bien definidos producirán una buena coincidencia de los puntos de los datos con la curva analítica. Use esta nueva relación y grafique otra vez los datos del esfuerzo *versus* los de tensión, y esta curva analítica nueva.

20.17 El espesor de la retina cambia durante ciertas enfermedades oculares. Una forma de medir dicho espesor es proyectar un láser de energía muy baja hacia la retina y grabar las reflexiones en una película. Debido a las propiedades ópticas del ojo, las reflexiones de la superficie frontal y trasera de la retina aparecerán en la película como dos líneas separadas por cierta distancia. Esta distancia es proporcional al espesor de la retina. Los datos siguientes se tomaron de una película grabada. Ajuste a los datos dos curvas con forma gaussiana de altura y ubicación arbitrarias, y determine la distancia entre los centros de los dos picos. Una curva gaussiana tiene la forma

$$f(x)\frac{ke^{-k^2(x-a)^2}}{\sqrt{\pi}}$$

donde k y a son constantes que relacionan la altura con el centro del pico, respectivamente.

Posición	Intensidad de luz	Posición	Intensidad de luz
0.17	5.10	0.31	25.31
0.18	5.10	0.32	23.79
0.19	5.20	0.33	18.44
0.20	5.87	0.34	12.45
0.21	8.72	0.35	8.22
0.22	16.04	0.36	6.12
0.23	26.35	0.37	5.35
0.24	31.63	0.38	5.15
0.25	26.51	0.39	5.10
0.26	16.68	0.40	5.10
0.27	10.80	0.41	5.09
0.28	11.26	0.42	5.09
0.29	16.05	0.43	5.09
0.30	21.96	0.44	5.09

20.18 Los datos que se tabulan en seguida se generaron en un experimento con contenido inicial de cianato de amonio puro (NH_4OCN). Se sabe que tales cambios de concentración se pueden modelar por la siguiente ecuación:

$$c = \frac{c_0}{1 + kc_0t}$$

donde c_0 y k son parámetros. Use una transformación para linealizar esta ecuación. Luego use regresión lineal para predecir la concentración a t = 160 min.

t (min)	0	20	50	65	150
c (moles/L)	0.381	0.264	0.180	0.151	0.086

Ingeniería civil/ambiental

20.19 A continuación se enlistan los esfuerzos cortantes, en kilopascales (kPa), de nueve especímenes tomados a distintas profundidades de un estrato arcilloso. Estime el esfuerzo cortante a la profundidad de 4.5 m.

Profundidad, m	1.9	3.1	4.2	5.1	5.8	6.9	8.1	9.3	10.0
Esfuerzo, kPa	14.4	28.7	19.2	43.1	33.5	52.7	71.8	62.2	76.6

20.20 Se realizó un estudio de ingeniería del transporte para determinar el diseño apropiado de pistas para bicicletas. Se recabaron datos del ancho de las pistas y la distancia promedio entre las bicicletas y los autos en circulación. Los datos de 9 calles son

Tabla P20.21 Concentración de oxígeno disuelto en agua como función de la temperatura (°C) y la concentración de cloruro (g/L).

	Oxígeno disuelto (mg/L) para la temperatura (°C) y la concentración de cloruro (g/L)		
T, °C	***c* = 0 g/L**	***c* = 10 g/L**	***c* = 20 g/L**
0	14.6	12.9	11.4
5	12.8	11.3	10.3
10	11.3	10.1	8.96
15	10.1	9.03	8.08
20	9.09	8.17	7.35
25	8.26	7.46	6.73
30	7.56	6.85	6.20

Distancia, m	2.4	1.5	2.4	1.8	1.8	2.9	1.2	3	1.2
Ancho de la pista, m	2.9	2.1	2.3	2.1	1.8	2.7	1.5	2.9	1.5

a) Grafique los datos.
b) Ajuste una línea recta a los datos con regresión lineal. Agregue esta línea a la gráfica.
c) Si se considera que la distancia promedio mínima de seguridad entre las bicicletas y los autos en circulación es de 2 m, determine el ancho de pista mínimo correspondiente.

20.21 En la tabla P20.21 se enlista la concentración de saturación del oxígeno disuelto en agua como función de la temperatura y la concentración de cloruro. Utilice interpolación para estimar el nivel de oxígeno disuelto para $T = 18°C$ con cloruro = 10 g/L.

20.22 Para los datos de la tabla P20.21, use regresión polinomial para obtener una ecuación predictiva de tercer grado para la concentración del oxígeno disuelto como función de la temperatura, para el caso en que la concentración de cloruro es igual a 10 g/L. Emplee la ecuación para estimar la concentración de oxígeno disuelto para $T = 8°C$.

20.23 Use regresión lineal múltiple para obtener una ecuación predictiva para la concentración del oxígeno disuelto como función de la temperatura y el cloruro, con base en los datos de la tabla P20.21. Use la ecuación para estimar la concentración de oxígeno disuelto para una concentración de cloruro de 5 g/L en $T = 17°C$.

20.24 En comparación con los modelos de los problemas 20.22 y 20.23, es posible plantear la hipótesis de un modelo algo más elaborado que toma en cuenta el efecto tanto de la temperatura como del cloruro sobre la saturación del oxígeno disuelto, el cual tiene la forma siguiente:

$$o_s = a_0 + f_3(T) + f_1(c)$$

Es decir, una constante más un polinomio de tercer grado en la temperatura y una relación lineal en el cloruro, se supone que dan resultados mejores. Use el enfoque lineal general de mínimos cuadrados para ajustar este modelo a los datos de la tabla P20.21. Emplee la ecuación resultante para estimar la concentración de oxígeno disuelto para una concentración de cloruro de 10 g/L a $T = 20°C$.

20.25 En la ingeniería de recursos hidráulicos, el tamaño de los almacenamientos depende de estimaciones exactas del flujo de agua en el río que se va a captar. Para ciertos ríos es difícil obtener registros históricos extensos de dichos datos de flujo. Por el contrario, es frecuente que se disponga de datos meteorológicos sobre la precipitación que se extienden mucho hacia el pasado. Por lo tanto, con frecuencia resulta útil determinar una relación entre el flujo y la precipitación. Entonces, esta relación se utiliza para estimar los flujos durante los años en que sólo se dispone de medidas pluviales. Se dispone de los datos siguientes para un río que va a represarse:

Precipitación, cm	88.9	108.5	104.1	139.7	127	94	116.8	99.1
Flujo, m^3/s	14.6	16.7	15.3	23.2	19.5	16.1	18.1	16.6

a) Grafique los datos.
b) Ajuste una línea recta a los datos por medio de regresión lineal. Sobreponga esta línea a su gráfica.
c) Use la línea de mejor ajuste para predecir el flujo anual de agua si la precipitación es de 120 cm.
d) Si el área de drenaje es de 1100 km^2, estime la fracción de la precipitación que se pierde a través de procesos como la evaporación, infiltración y uso consuntivo.

20.26 La concentración del fósforo total (p en mg/m^3) y clorofila a (c en mg/m^3) para cada uno de los Grandes Lagos en el año de 1970, fue

	p	***c***
Lago Superior	4.5	0.8
Lago Michigan	8.0	2.0
Lago Hurón	5.5	1.2
Lago Erie:		
Cuenca oeste	39.0	11.0
Cuenca central	19.5	4.4
Cuenca este	17.5	3.8
Lago Ontario	21.0	5.5

La concentración de clorofila a indica cuánta vida vegetal se encuentra en suspensión en el agua. Al ser así, indica la claridad y visibilidad del agua. Use los datos anteriores para determinar la relación de c como función de p. Emplee la ecuación para predecir el nivel de clorofila que puede esperarse si se utiliza el tratamiento del agua para abatir a 10 mg/m^3 la concentración de fósforo del Lago Erie occidental.

20.27 El esfuerzo vertical σ_z bajo la esquina de un área rectangular sujeta a una carga uniforme de intensidad q, está dada por la solución de la ecuación de Boussinesq:

$$\sigma = \frac{q}{4\pi}\left[\frac{2mn\sqrt{m^2+n^2+1}}{m^2+n^2+1+m^2n^2}\frac{m^2+n^2+2}{m^2+n^2+1}\right.$$
$$\left.+\text{sen}^{-1}\left(\frac{2mn\sqrt{m^2+n^2+1}}{m^2+n^2+1+m^2n^2}\right)\right]$$

Debido a que es inconveniente resolver esta ecuación manualmente, ha sido reformulada como

$$\sigma_z = qf_z(m, n)$$

donde $f_z(m, n)$ se denomina el valor de influencia, y m y n son razones adimensionales, con $m = a/z$ y $n = b/z$, y a y b se encuentran definidas en la figura P20.27. Después se tabula el valor de influencia, una parte de la cual está dada en la tabla P20.27. Si $a = 4.6$ y $b = 14$, use un polinomio de interpolación de tercer grado para calcular σ_z a una profundidad de 10 m por debajo de la esquina de una cornisa rectangular que está sujeta a una carga total de 100 t (toneladas métricas). Exprese su respuesta en toneladas por metro cuadrado. Observe que q es igual a la carga por área.

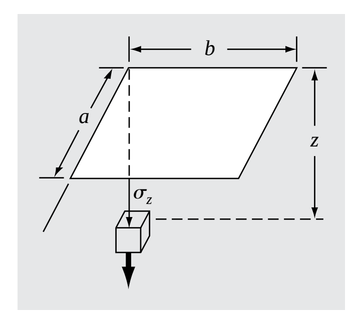


Figura P20.27

Tabla P20.27

m	*n* = 1.2	*n* = 1.4	*n* = 1.6
0.1	0.02926	0.03007	0.03058
0.2	0.05733	0.05894	0.05994
0.3	0.08323	0.08561	0.08709
0.4	0.10631	0.10941	0.11135
0.5	0.12626	0.13003	0.13241
0.6	0.14309	0.14749	0.15027
0.7	0.15703	0.16199	0.16515
0.8	0.16843	0.17389	0.17739

20.28 Tres organismos patógenos decaen en forma exponencial en aguas de un lago de acuerdo con el modelo siguiente:

$$p(t) = Ae^{-1.5t} + Be^{-0.3t} + Ce^{-0.05t}$$

Estime la población inicial de cada organismo (A, B y C), dadas las mediciones siguientes:

t, h	0.5	1	2	3	4	5	6	7	9
p, (*t*)	6.0	4.4	3.2	2.7	2.2	1.9	1.7	1.4	1.1

20.29 El mástil de un velero tiene un área de sección transversal de 10.65 cm², y está construido de una aleación experimental de aluminio. Se llevaron a cabo pruebas para definir la relación entre el esfuerzo y la tensión. Los resultados de las pruebas fueron los que siguen:

Tensión, cm/cm	0.0032	0.0045	0.0055	0.0016	0.0085	0.0005
Esfuerzo, N/cm²	4 970	5 170	5 500	3 590	6 900	1 240

Los esfuerzos ocasionados por el viento se calculan como F/A_c; donde F = fuerza en el mástil y A_c = área de la sección transversal del mástil. Después, este valor se sustituye en la ley de Hooke para determinar la deflexión del mástil, ΔL = tensión × L, donde L = longitud del mástil. Si la fuerza del viento es de 25 000 N, use los datos para estimar la deflexión de un mástil de 9 m.

20.30 En la ingeniería ambiental, las reacciones enzimáticas se utilizan mucho para caracterizar reacciones mediadas biológicamente. A continuación se dan expresiones de tasas propuestas para una reacción enzimática, donde $[S]$ es la concentración del sustrato y v_0 es la tasa inicial de la reacción. ¿Qué fórmula se ajusta mejor a los datos experimentales?

$$v_0 = k[S] \quad v_0 = \frac{k[S]}{K+[S]} \quad v_0 = \frac{k[S]^2}{K+[S]^2} \quad v_0 = \frac{k[S]^3}{K+[S]^3}$$

[*S*], M	**Tasa inicial, 10⁻⁶ M/s**
0.01	6.3636×10^{-5}
0.05	7.9520×10^{-3}
0.1	6.3472×10^{-2}
0.5	6.0049
1	17.690
5	24.425
10	24.491
50	24.500
100	24.500

20.31 Los ingenieros ambientales que estudian los efectos de la lluvia ácida deben determinar el valor del producto iónico del agua K_w como función de la temperatura. Los científicos sugieren la ecuación siguiente para modelar dicha relación.

$$-\log_{10} K_w = \frac{a}{T_a} + b\log_{10} T_a + cT_a + d$$

donde T_a es la temperatura absoluta (K), y a, b, c y d son parámetros. Emplee los datos siguientes y la regresión para estimar los parámetros:

T_a (*K*)	273.15	283.15	293.15
K_w	1.164×10^{-15}	2.950×10^{-15}	6.846×10^{-15}

T_a (*K*)	303.15	313.15
K_w	1.467×10^{-14}	2.929×10^{-14}

20.32 Un ingeniero ambiental reportó los datos tabulados más abajo de un experimento para determinar la tasa de crecimiento de bacterias,

k, en función de la concentración de oxígeno, c. Se sabe que a dichos datos se les puede aplicar el modelo de la siguiente ecuación:

$$k = \frac{k_{\text{máx}} c^2}{c_s + c^2}$$

donde c_s y $k_{\text{máx}}$ son parámetros. Emplee una transformación para linealizar esta ecuación, y después use regresión lineal para estimar c_s y $k_{\text{máx}}$ y predecir la tasa de crecimiento en $c = 2$ mg/L.

c (mg/L)	0.5	0.8	1.5	2.5	4
k (por día)	1.1	2.4	5.3	7.6	8.9

20.33 El siguiente modelo se usa frecuentemente en ingeniería ambiental para parametrizar el efecto de la temperatura, T (°C), en las rapideces de reacción bioquímica, k (por día),

$$k = k_{20}\theta^{T-20}$$

donde k_{20} y θ son parámetros. Use una transformación para linealizar esta ecuación. Luego use regresión lineal para estimar k_{20} y θ y prediga la rapidez de reacción a T = 17°C.

T (°C)	6	12	18	24	30
k (por día)	0.14	0.20	0.31	0.46	0.69

20.34 Como miembro de Ingenieros sin Fronteras, usted está trabajando en una comunidad que tiene su agua de beber contaminada. En t = 0, usted agrega un desinfectante a una cisterna que está contaminada con bacterias. Usted hace las siguientes mediciones a distintos tiempos desde ese momento:

t (h)	2	4	6	8	10
c (#/100 mL)	430	190	80	35	16

Si el agua es segura para beber cuando la concentración caiga por debajo de 5 #/100 mL, estime el tiempo en que la concentración caerá por debajo de este límite.

Ingeniería eléctrica

20.35 Lleve a cabo los mismos cálculos que en la sección 20.3, pero analice los datos generados con $f(t) = 4 \cos(5t) - 7 \text{ sen}(3t) + 6$.

20.36 Se mide la caída de voltaje V a través de un resistor para cierto número de valores distintos de corriente i. Los resultados son

i	0.25	0.75	1.25	1.5	2.0
V	–0.45	–0.6	0.70	1.88	6.0

Utilice interpolación de polinomios de primero a cuarto grados para estimar la caída de voltaje para i = 1.15. Interprete los resultados.

20.37 Repita el cálculo para el problema 20.36, pero use regresión polinomial para obtener ecuaciones de mejor ajuste de grados 1 a 4 con el uso de todos los datos. Grafique y evalúe sus resultados.

20.38 Se mide con gran precisión la corriente en un conductor como función del tiempo:

t	0	0.1250	0.2500	0.3750	0.5000
i	0	6.24	7.75	4.85	0.0000

Determine el valor de i en t = 0.23.

20.39 Los datos siguientes se tomaron de un experimento para medir la corriente en un conductor para varios voltajes aplicados:

V, V	2	3	4	5	7	10
i, A	5.2	7.8	10.7	13	19.3	27.5

a) Sobre la base de una regresión lineal de estos datos, determine la corriente para un voltaje de 3.5 V. Grafique la línea y los datos, y evalúe el ajuste.

b) Repita la regresión y fuerce la intersección para que sea cero.

20.40 Se sabe que la caída de voltaje a través de un inductor sigue la ley de Faraday:

$$V_L = L\frac{di}{dt}$$

donde V_L es la caída del voltaje (en volts), L es la inductancia (en henrys; 1 H = 1 V · s/A) e i es la corriente (en amperes). Emplee los datos siguientes para estimar L:

di/dt, A/s	1	2	4	6	8	10
V_L, V	5.5	12.5	17.5	32	38	49

¿Cuál es el significado, si hubiera alguno, de la intersección de la ecuación de regresión que se obtiene con estos datos?

20.41 La ley de Ohm establece que la caída de voltaje V a través de un resistor ideal es linealmente proporcional a la corriente i que fluye a través del resistor, como en $V = iR$, donde R es la resistencia. Sin embargo, los resistores reales no siempre obedecen a la ley de Ohm. Suponga usted que lleva a cabo algunos experimentos muy precisos para medir la caída de voltaje y la corriente correspondiente para un resistor. Los resultados, que se enlistan en la tabla P20.41, sugieren una relación curvilínea, más que la línea recta que representa la ley de Ohm. A fin de cuantificar dicha relación debe ajustarse una curva a los datos. Debido al error en la medición, es común que la regresión sea el método preferido de ajuste de curvas para analizar dichos datos experimentales. Sin embargo, la suavidad de la relación, así como la precisión de los métodos experimentales, sugieren que quizá sería apropiada la interpolación. Utilice la interpolación de polinomios de Newton para ajustar los datos y calcular V para i = 0.10. ¿Cuál es el grado del polinomio que se usó para generar los datos?

Tabla P20.41 Datos experimentales para la caída del voltaje a través de un resistor sujeto a distintos niveles de corriente.

i	–2	–1	–0.5	0.5	1	2
V	–637	–96.5	–20.5	20.5	96.5	637

20.42 Repita el problema 20.41, pero determine los coeficientes del polinomio (véase la sección 18.4) que se ajusta a los datos de la tabla P20.41.

20.43 Se realiza un experimento para determinar la elongación porcentual de un material conductor de electricidad como función de la temperatura. Los datos que resultan se presentan en seguida. Prediga la elongación porcentual para una temperatura de 400°C.

Temperatura, °C	200	250	300	375	425	475	600
% de elongación	7.5	8.6	8.7	10	11.3	12.7	15.3

20.44 Es frecuente que en los análisis avanzados de ingeniería surjan funciones de Bessel, como en el estudio de campos eléctricos. Dichas funciones por lo general no son susceptibles de evaluarse en forma directa y, por ello, no es raro que estén compiladas en tablas matemáticas estándar. Por ejemplo,

x	1.8	2	2.2	2.4	2.6
$J_1(x)$	0.5815	0.5767	0.556	0.5202	0.4708

Estime $J_1(2.1)$, *a*) con el uso de un polinomio de interpolación y *b*) con trazadores cúbicos. Observe que el valor verdadero es 0.568292.

20.45 La población (p) de una comunidad pequeña en los suburbios de una ciudad crece con rapidez durante un periodo de 20 años:

t	0	5	10	15	20
p	100	200	450	950	2 000

Como ingeniero que trabaja para una compañía de infraestructura, el lector debe pronosticar la población que habrá dentro de 5 años a fin de anticipar la demanda de energía. Emplee un modelo exponencial y regresión lineal para efectuar dicha predicción.

Ingeniería mecánica/aeroespacial

20.46 Con base en la tabla 20.4, utilice interpolación lineal y cuadrática para calcular el valor de Q para $D = 1.23$ ft y $S = 0.001$ ft/ft. Compare sus resultados con el mismo valor calculado con la fórmula que se obtuvo en la sección 20.4.

20.47 Reproduzca la sección 20.4, pero desarrolle una ecuación para predecir la pendiente como función del diámetro y flujo. Compare sus resultados con los de la fórmula de la sección 20.4 y analice su respuesta.

20.48 La viscosidad dinámica del agua $\mu(10^{-3}\ \text{N} \cdot \text{s/m}^2)$ se relaciona con la temperatura T(°C), de la manera siguiente:

T	0	5	10	20	30	40
μ	1.787	1.519	1.307	1.002	0.7975	0.6529

a) Grafique los datos.
b) Use interpolación para predecir μ con $T = 7.5$°C.
c) Emplee regresión polinomial para ajustar una parábola a los datos a fin de hacer la misma predicción.

20.49 La *Ley de Hooke*, que se cumple cuando un resorte no se estira más allá de cierto límite, significa que la extensión de este resorte y la fuerza que se le aplica están relacionadas linealmente. La proporcionalidad está parametrizada por la constante k del resorte. Un valor para dicho parámetro se establece en forma experimental con la colocación de pesos conocidos en el resorte y la medición de la compresión que resulta. Tales datos aparecen en la tabla P20.49 y

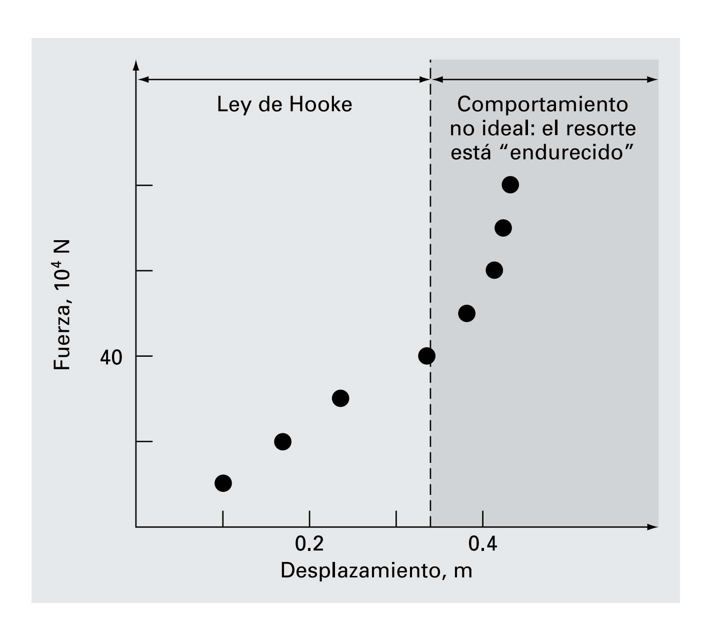


Figura P20.49
Gráfica de la fuerza (en 10^4 newtons) *versus* el desplazamiento (en metros) para el resorte del sistema de suspensión del automóvil.

están graficados en la figura P20.49. Observe que por arriba de un peso de 40×10^4 N, la relación lineal entre la fuerza y el desplazamiento desaparece. Esta clase de comportamiento es común de lo que se denomina "resorte endurecido". Emplee regresión lineal para determinar un valor de k para la parte lineal de este sistema. Además, ajuste una relación no lineal a la parte no lineal.

20.50 Repita el problema 20.49 pero ajuste una curva de potencias a todos los datos de la tabla P20.49. Comente sus resultados.

20.51 La distancia que se requiere para detener un automóvil consiste en componentes tanto de pensamiento como de frenado, cada una de las cuales es función de la velocidad. Se recabaron los siguientes datos experimentales para cuantificar dicha relación. Desarrolle la ecuación de mejor ajuste para ambos componentes, pensamiento y frenado. Utilice estas ecuaciones para estimar la distancia total en que se detiene un auto que viaja a 110 km/h.

Velocidad, km/h	30	45	60	75	90	120
Pensamiento, m	5.6	8.5	11.1	14.5	16.7	22.4
Frenado, m	5.0	12.3	21.0	32.9	47.6	84.7

20.52 Se realiza un experimento para definir la relación entre el esfuerzo aplicado y el tiempo para que se fracture cierto tipo de acero inoxidable. Se aplican ocho valores distintos de esfuerzo, y los datos resultantes son

Esfuerzo aplicado, x, kg/mm^2	5	10	15	20	25	30	35	40
Tiempo para la fractura, y, h	40	30	25	40	18	20	22	15

Tabla P20.49 Valores experimentales para la elongación x y la fuerza F para el resorte de un sistema de suspensión de automóvil

Desplazamiento, m	0.10	0.17	0.27	0.35	0.39	0.42	0.43	0.44
Fuerza, 10^4N	10	20	30	40	50	60	70	80

Grafique los datos y después desarrolle la ecuación de mejor ajuste para predecir el tiempo de fractura para un esfuerzo aplicado de 20 kg/mm^2.

20.53 La aceleración debida a la gravedad a una altitud y por encima de la superficie de la Tierra está dada por

y, m	0	30 000	60 000	90 000	120 000
g, m/s^2	9.8100	9.7487	9.6879	9.6278	9.5682

Calcule g para y = 55 000 m.

20.54 De un procedimiento de prueba se obtuvieron la tasa de arrastre $\dot{\varepsilon}$ que es la tasa de tiempo a que aumenta la tensión, y de esfuerzos, los cuales se presentan a continuación. Con el uso de una ley de curva de potencias para ajustar,

$$\dot{\varepsilon} = B\sigma^m$$

encuentre el valor de B y m. Grafique sus resultados con el empleo de una escala log-log.

Tasa de arrastre, min^{-1}	0.0004	0.0011	0.0021	0.0031
Esfuerzo, MPa	5.775	8.577	10.874	12.555

20.55 Al examinar el comportamiento viscoso de un fluido es práctica común graficar la tasa de corte (gradiente de velocidad)

$$\frac{dv}{dy} = \dot{\gamma}$$

en las abscisas *versus* el esfuerzo cortante (τ) en las ordenadas. Cuando un fluido muestra un comportamiento en línea recta entre esas dos variables, se denomina *fluido newtoniano*, y la relación resultante es

$$\tau = \mu\dot{\gamma}$$

donde μ es la viscosidad del fluido. Muchos fluidos comunes siguen este comportamiento como el agua, leche y aceite. Los fluidos que no se comportan de esa manera, se llaman *no newtonianos*. En la figura P20.55 se muestran algunos ejemplos de fluidos no newtonianos.

Para *plásticos Bingham*, hay un esfuerzo inducido τ_y que debe superarse para que el flujo comience,

$$\tau = \tau_y + \mu\dot{\gamma}$$

Un ejemplo común es la pasta de dientes.

Para fluidos *pseudoplásticos o "fluidificados por esfuerzo cortante"*, el esfuerzo cortante se eleva a la potencia n menor que uno,

$$\tau = \mu\dot{\gamma}^n$$

Fluidos como el yogurt, la mayonesa y el champú experimentan una disminución de viscosidad al aumentar el esfuerzo. Observe que para los casos en que $n > 1$, que se llaman fluidos *dilatantes* (o *"espesados por esfuerzo cortante"*), la viscosidad efectivamente aumenta al incrementar el esfuerzo. Los ejemplos incluyen almidón en agua y arena de playa mojada.

Los siguientes datos muestran la relación entre el esfuerzo cortante τ y la tasa de tensión cortante τ_γ es la cantidad de esfuerzo que se debe exceder antes de que comience el flujo. Encuentre la viscosidad μ (pendiente), τ_γ y el valor de r^2 usando un método de regresión. ¿Cuál es el tipo de fluido?

Esfuerzo τ, N/m^2	3.25	4.25	4.65	5.65	6.05
Tasa de tensión cortante, $\dot{\gamma}$, 1/s	0.9	2.1	2.9	4.1	4.9

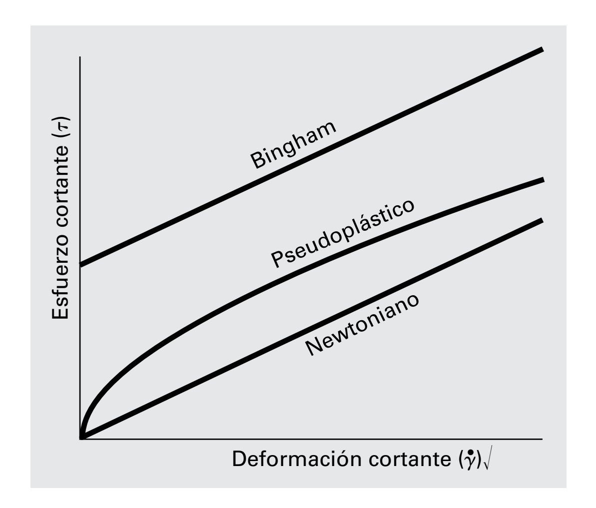


Figura P20.55

20.56 La relación entre el esfuerzo τ y la tasa de deformación cortante $\dot{\gamma}$ para un fluido pseudoplástico (véase el problema 20.55), puede expresarse con la ecuación $\tau = \mu\dot{\gamma}^n$. Los siguientes datos provienen de una solución al 5% de hidroxetilcelulosa en agua. Usando un ajuste por ley de potencias, encuentre los valores de μ y n.

Tasa de tensión cortante, $\dot{\gamma}$, 1/s	50	70	90	110	130
Esfuerzo τ, N/m^2	6.01	7.48	8.59	9.19	10.21

20.57 Se mide la velocidad u del aire que fluye a varias distancias y de una superficie plana. Ajuste una curva a esos datos si se supone que la velocidad en la superficie es igual a cero (y = 0). Utilice su resultado para determinar el esfuerzo cortante ($\mu\, du/dy$) en la superficie. ($\mu = 1.8 \times 10^{-5}$ N · s/m^2)

y, m	0.002	0.006	0.012	0.018	0.024
u, m/s	0.287	0.899	1.915	3.048	4.299

20.58 La *ecuación de Andrade* ha sido propuesta como modelo del efecto de la temperatura sobre la viscosidad,

$$\mu = De^{B/Ta}$$

donde μ = viscosidad dinámica del agua (10^{-3} N · s/m^2), T_a = temperatura absoluta (K), y D y B son parámetros. Ajuste este modelo a los datos del agua del problema 20.48.

20.59 Desarrolle ecuaciones para ajustar los calores específicos ideales c_p (kJ/kg · K), como función de la temperatura T (K), para varios gases, según se enlistan en la tabla P20.59.

20.60 Se mide la temperatura en varios puntos de una placa calentada (véase la tabla P20.60). Estime la temperatura en *a*) x = 4, y = 3.2 y *b*) x = 4.3, y = 2.7.

Tabla P20.59 Calores específicos ideales, c_p (kJ/kg · K) como función de la temperatura para distintos gases.

Gas	250 K	300 K	350 K	450 K	550 K	650 K	800 K	900 K	1000 K
H_2	14.051	14.307	14.427	14.501	14.53	14.571	14.695	14.822	14.983
CO_2	0.791	0.846	0.895	0.978	1.046	1.102	1.169	1.204	1.234
O_2	0.913	0.918	0.928	0.956	0.988	1.017	1.054	1.074	1.09
N_2	1.039	1.039	1.041	1.049	1.065	1.086	1.121	1.145	1.167

Tabla P20.60 Temperaturas (°C) en varios puntos de una placa cuadrada calentada.

	***x* = 0**	***x* = 2**	***x* = 4**	***x* = 6**	***x* = 8**
***y* = 0**	100.00	90.00	80.00	70.00	60.00
***y* = 2**	85.00	64.49	53.50	48.15	50.00
***y* = 4**	70.00	48.90	38.43	35.03	40.00
***y* = 6**	55.00	38.78	30.39	27.07	30.00
***y* = 8**	40.00	35.00	30.00	25.00	20.00

20.61 Los datos de la página siguiente se obtuvieron de una prueba de arrastre que se llevó a cabo a la temperatura ambiente sobre un alambre compuesto de 40% de hojalata, 60% de plomo y un núcleo sólido soldado. Esto se realizó por medio de la medición del incremento en la tensión durante el tiempo mientras se aplicaba una carga constante a un espécimen de prueba.

Con un método de regresión lineal, encuentre *a*) la ecuación de la línea que mejor ajuste los datos y *b*) el valor de r^2. Grafique sus resultados. ¿La línea que pasa por el origen —es decir, en el tiempo cero— debe tener alguna tensión? Si la línea no pasa por el origen, fuércela a hacerlo. ¿La nueva recta representa la tendencia de los datos? Sugiera una ecuación nueva que satisfaga la tensión igual a cero en el tiempo igual a cero, y también represente la tendencia de los datos.

Tiempo, min	Tensión, %	Tiempo, min	Tensión, %	Tiempo, min	Tensión, %
0.085	0.10	3.589	0.26	7.092	0.43
0.586	0.13	4.089	0.30	7.592	0.45
1.086	0.16	4.590	0.32	8.093	0.47
1.587	0.18	5.090	0.34	8.593	0.50
2.087	0.20	5.591	0.37	9.094	0.52
2.588	0.23	6.091	0.39	9.594	0.54
3.088	0.25	6.592	0.41	10.097	0.56

EPÍLOGO: PARTE CINCO

PT5.4 ALTERNATIVAS

La tabla PT5.4 ofrece un resumen de las ventajas y las desventajas de los métodos para el ajuste de curvas. Las técnicas se dividen en dos grandes categorías, según sea la incertidumbre de los datos. Para mediciones imprecisas la regresión se utiliza para desarrollar una curva que "mejor" se ajuste a la tendencia global de los datos, sin que necesariamente pase a través de alguno de los puntos. Para mediciones precisas se usa la interpolación para desarrollar una curva que pase justo a través de los puntos.

Todos los métodos de regresión están diseñados para ajustar funciones que minimicen la suma de los cuadrados de los residuos entre los datos y la función. Tales métodos se denominan de regresión por mínimos cuadrados. La regresión lineal por mínimos cuadrados se usa para casos donde una variable dependiente y otra independiente se relacionan entre sí en forma lineal. Para situaciones donde una variable dependiente y una independiente exhiben un comportamiento curvilíneo, hay varias opciones disponibles. En algunos casos, se emplean transformaciones para linealizar el comportamiento. En estos casos, se aplica una regresión lineal a las variables transformadas con el propósito de determinar la mejor línea recta. De manera alternativa, la regresión polinomial se utiliza para ajustar una curva directamente a los datos.

La regresión lineal múltiple se utiliza cuando una variable dependiente es una función lineal de dos o más variables independientes. Las transformaciones logarítmicas también se aplican a este tipo de regresión en aquellos casos donde la dependencia múltiple es curvilínea.

La regresión polinomial y la lineal múltiple (observe que la regresión lineal simple es una particularidad de ambas) pertenecen a una clase más general de modelos de mínimos cuadrados lineales. Se clasifican de esta manera porque son lineales respecto a sus coeficientes. Por lo común estos modelos se implementan a través de la solución de sistemas algebraicos lineales, que algunas veces están mal condicionados. Sin embargo, en muchos problemas de ingeniería (se tienen ajustes de grado inferior), afortunadamente, no ocurre. En los casos donde esto represente un problema se cuenta con algunos procedimientos alternativos. Por ejemplo, existe una técnica llamada de polinomios ortogonales, para realizar la regresión polinomial (véase la sección PT5.6).

Las ecuaciones que no son lineales respecto a sus coeficientes se denominan no lineales. Hay técnicas de regresión especiales para ajustar tales ecuaciones. Éstos son métodos aproximados que

Tabla PT5.4 Comparación de las características de los diferentes métodos para el ajuste de curvas.

Método	Error asociado con datos	Coincidencia con los datos individuales	Núm. de puntos que coinciden exactamente	Dificultad de programación	Comentarios
Regresión					
Regresión lineal	Grande	Aproximada	0	Fácil	
Regresión polinomial	Grande	Aproximada	0	Moderada	El error de redondeo se vuelve pronunciado en versiones de grado superior
Regresión lineal múltiple	Grande	Aproximada	0	Moderada	
Regresión no lineal	Grande	Aproximada	0	Difícil	
Interpolación					
Polinomios de Newton en diferencias divididas	Pequeña	Exacta	$n + 1$	Fácil	Se prefiere para análisis exploratorios
Polinomios de Lagrange	Pequeña	Exacta	$n + 1$	Fácil	Se prefiere cuando se conoce el grado
Trazadores cúbicos	Pequeña	Exacta	Ajuste por segmentos a los puntos asociados con datos	Moderada	Primera y segunda derivadas iguales en nodos

empiezan con un parámetro inicial estimado y después, iterativamente, llegan a valores que minimizan la suma de los cuadrados.

La interpolación polinomial está diseñada para ajustar un único polinomio de n-ésimo grado que pasa exactamente a través de los $n + 1$ puntos que se tienen como datos. Este polinomio se presenta en dos formas alternativas. La interpolación polinomial de Newton en diferencias divididas es ideal en aquellos casos donde se conoce el grado del polinomio. El polinomio de Newton resulta apropiado en tales situaciones, ya que se programa en forma sencilla en un formato que sirve para comparar resultados con diferentes grados. Además, un error estimado simplemente se puede incorporar en la técnica. Así, usted puede comparar y elegir de los resultados usando varios polinomios de diferente grado.

La interpolación de polinomios de Lagrange es una forma alternativa que es conveniente cuando el grado se conoce de antemano. En dichas situaciones, la versión de Lagrange es más fácil de programar y no requiere del cálculo ni el almacenamiento de diferencias finitas divididas.

Otro procedimiento para ajustar curvas es la interpolación mediante trazadores. Esta técnica ajusta un polinomio de grado inferior para cada intervalo entre los puntos dados. El ajuste se suaviza igualando las derivadas de polinomios adyacentes al mismo valor en sus puntos de unión. Los trazadores cúbicos son el modelo más común. Los trazadores son de gran utilidad cuando se ajustan a datos que por lo general son suaves; pero que exhiben áreas locales de cambio abrupto. Tales datos tienden a inducir oscilaciones desordenadas cuando se interpolan polinomios de grado superior. Los trazadores cúbicos son menos propensos a esas oscilaciones debido a que están limitados a variaciones de tercer grado.

Más allá del caso unidimensional, se puede implementar interpolación para datos multidimensionales. Tanto los polinomios de interpolación como los trazadores se pueden usar para este propósito. Hay paquetes de software disponibles para agilizar tales aplicaciones.

El último método que se estudia en esta parte del libro es la aproximación de Fourier, la cual trata con el uso de funciones trigonométricas para aproximar diversas formas de ondas. En contraste con las otras técnicas, el mayor énfasis de este procedimiento no es ajustar una curva a los datos; sino que el ajuste de la curva se emplee para analizar las frecuencias características de una señal. En particular, la transformada rápida de Fourier permite transformar eficientemente una función del dominio del tiempo al de la frecuencia, para entender su estructura armónica.

PT5.5 RELACIONES Y FÓRMULAS IMPORTANTES

La tabla PT5.5 resume información importante que se presentó en la parte cinco. Esta tabla se puede consultar para tener un rápido acceso a las relaciones y las fórmulas importantes.

PT5.6 MÉTODOS AVANZADOS Y REFERENCIAS ADICIONALES

Aunque la regresión polinomial con ecuaciones normales es adecuada para muchos problemas de ingeniería, hay contextos de problemas donde su sensibilidad a los errores de redondeo presenta serias limitaciones. Un procedimiento alternativo basado en *polinomios ortogonales* puede disminuir esos efectos. Deberá observarse que este procedimiento no da una ecuación de mejor ajuste, sino más bien predicciones individuales para valores dados de la variable independiente. Se recomienda consultar a Shampine y Allen (1973) y Guest (1961) para mayor información acerca de los polinomios ortogonales.

Mientras que la técnica de polinomios ortogonales es útil para desarrollar una regresión polinomial, no representa una solución al problema de inestabilidad para el modelo de regresión lineal general [ecuación (17.23)]. Hay un procedimiento alternativo basado en la *descomposición de valores singulares*, llamado método SVD, para dicho propósito. Información sobre este procedimiento se encuentra en Forsythe y colaboradores (1977), Lawson y Hanson (1974), y Press y colaboradores (1992).

Además del algoritmo de Gauss-Newton, existen varios métodos de optimización que se utilizan de manera directa con la finalidad de desarrollar un ajuste por mínimos cuadrados para una ecuación no lineal. Dichas técnicas de regresión no lineal incluyen los métodos de máximo descenso y de Marquardt (recuerde la parte cuatro). Para mayor información sobre regresión consulte a Draper y Smith (1981).

Tabla PT5.5 Resumen de la información importante presentada en la parte cinco.

Método	Formulación	Interpretación gráfica	Errores
Regresión lineal	$y = a_0 + a_1x$ donde $a_1 = \dfrac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$ $a_0 = \bar{y} - a_1\bar{x}$		$s_{y/x} = \sqrt{\dfrac{S_r}{n-2}}$ $r^2 = \dfrac{S_t - S_r}{S_t}$
Regresión polinomial	$y = a_0 + a_1x + \cdots + a_mx^m$ (Evaluación de las a equivalente a la solución de $m + 1$ ecuaciones algebraicas lineales)		$s_{y/x} = \sqrt{\dfrac{S_r}{n-(m+1)}}$ $r^2 = \dfrac{S_t - S_r}{S_t}$
Regresión lineal múltiple	$y = a_0 + a_1x_1 + \cdots + a_mx_m$ (Evaluación de las a equivalentes a la solución de $m + 1$ ecuaciones algebraicas lineales)		$s_{y/x} = \sqrt{\dfrac{S_r}{n-(m+1)}}$ $r^2 = \dfrac{S_t - S_r}{S_t}$
Interpolación polinomial de Newton en diferencias divididas*	$f_2(x) = b_0 + b_1(x - x_0) + b_2(x - x_0)(x - x_1)$ donde $b_0 = f(x_0)$ $b_1 = f[x_1, x_0]$ $b_2 = f[x_2, x_1, x_0]$		$R_2 = (x - x_0)(x - x_1)(x - x_2)\dfrac{f^{(3)}(\xi)}{6}$ o $R_2 = (x - x_0)(x - x_1)(x - x_2)f[x_3, x_2, x_1, x_0]$
Interpolación polinomial de Lagrange*	$f_2(x) = f(x_0)\left(\dfrac{x - x_1}{x_0 - x_1}\right)\left(\dfrac{x - x_2}{x_0 - x_2}\right) + f(x_1)\left(\dfrac{x - x_0}{x_1 - x_0}\right)\left(\dfrac{x - x_2}{x_1 - x_2}\right) + f(x_2)\left(\dfrac{x - x_0}{x_2 - x_0}\right)\left(\dfrac{x - x_1}{x_2 - x_1}\right)$		$R_2 = (x - x_0)(x - x_1)(x - x_2)\dfrac{f^{(3)}(\xi)}{6}$ o $R_2 = (x - x_0)(x - x_1)(x - x_2)f[x_3, x_2, x_1, x_0]$
Trazadores cúbicos	Una cúbica: $a_ix^3 + b_ix^2 + c_ix + d_i$ se ajusta a cada intervalo entre nodos. Primera y segunda derivadas son iguales en cada nodo	$a_1x^3 + b_1x^2 + c_1x + d_1$ nodo $a_2x^3 + b_2x^2 + c_2x + d_2$	

*Nota: Para simplificar, se muestran las versiones de segundo grado.

Todos los métodos de la parte cinco se estudiaron en términos de ajuste de curvas a datos. Además, quizá usted desee ajustar una curva a otra. El motivo principal de tal *aproximación funcional* es la representación de una función complicada mediante una versión más simple que sea más fácil de manipular. Una manera de hacerlo consiste en usar la función complicada para generar una tabla de valores. Después, las técnicas analizadas en esta parte del libro pueden usarse para ajustar polinomios a estos valores discretos.

Un procedimiento alternativo se basa en el *principio minimax* (véase la figura 17.2*c*). Este principio especifica que los coeficientes de la aproximación polinomial deben elegirse de tal forma que la discrepancia máxima sea lo más pequeña posible. Así, aunque la aproximación no sea tan buena como la que ofrece la serie de Taylor en el punto base, por lo general es mejor en todo el intervalo del ajuste. La *economización de Chebyshev* es un ejemplo de un procedimiento para una aproximación funcional basada en tal estrategia (Ralston y Rabinowitz, 1978; Gerald y Wheatley, 2004, y Carnahan, Luther y Wilkes, 1969).

Una alternativa importante en el ajuste de curvas es la combinación de trazadores con una regresión por mínimos cuadrados. Así, se genera un trazador cúbico de tal forma que no intercepte todos los puntos, pero que minimice la suma de los cuadrados de los residuos entre los datos y los trazadores. El procedimiento usa los denominados *trazadores B* como funciones base; se nombran así debido a su empleo como función *base*, y también por su forma de campana (bell) característica. Tales curvas son consistentes con un procedimiento de trazadores, puesto que la función y su primera y segunda derivadas serán continuas en los extremos. De esta forma se asegura la continuidad de $f(x)$ y sus derivadas en los nodos. Wold (1974), Prenter (1975) y Cheney y Kincaid (2008) ofrecen un análisis de tal procedimiento.

En resumen, con lo anterior se intenta proporcionarle alternativas para la exploración más profunda del tema. Asimismo, todas las referencias anteriores proporcionan descripciones de las técnicas básicas tratadas en la parte cinco. Le recomendamos consultar esas fuentes alternas para ampliar su comprensión de los métodos numéricos para el ajuste de curvas.

PARTE SEIS



DIFERENCIACIÓN E INTEGRACIÓN NUMÉRICAS

PT6.1 MOTIVACIÓN

El cálculo es la matemática del cambio. Como los ingenieros deben tratar en forma continua con sistemas y procesos que cambian, el cálculo es una herramienta esencial en nuestra profesión. En la esencia del cálculo están dos conceptos matemáticos relacionados: la diferenciación y la integración.

De acuerdo a la definición del diccionario, *diferenciar* significa "marcar por diferencias; distinguir;… percibir la diferencia en o entre". En el contexto de las matemáticas, la *derivada* sirve como el principal vehículo para la diferenciación, representa la razón de cambio de una variable dependiente con respecto a una variable independiente. Como se ilustra en la figura PT6.1, la definición matemática de la derivada empieza con una aproximación por diferencias:

$$\frac{\Delta y}{\Delta x} = \frac{f(x_i + \Delta x) - f(x_i)}{\Delta x} \tag{PT6.1}$$

donde y y $f(x)$ son representaciones alternativas de la variable dependiente y x es la variable independiente. Si se hace que Δx se aproxime a cero, como sucede en los movimientos mostrados desde la figura PT6.1*a* a la PT6.1*c*, el cociente de las diferencias se convierte en una derivada

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{f(x_i + \Delta x) - f(x_i)}{\Delta x}$$

donde dy/dx [que también se denota como y' o $f'(x_i)$] es la primera derivada de y con respecto a x evaluada en x_i. Como se observa en la descripción visual de la figura PT6.1*c*, la derivada evaluada es la pendiente de la recta tangente a la curva en x_i.

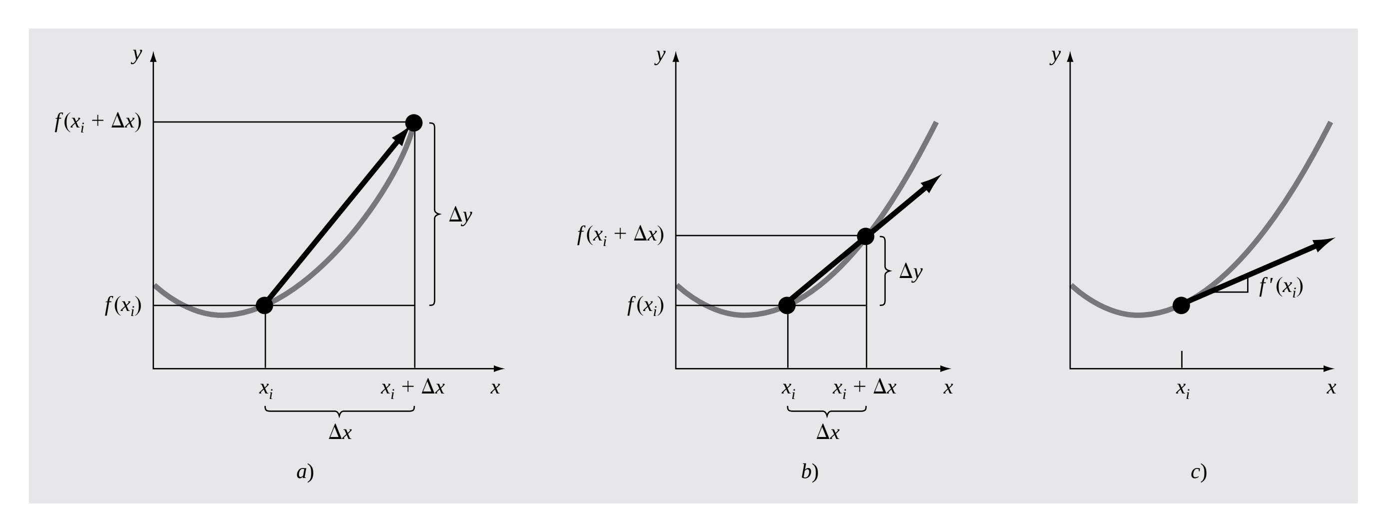


Figura PT6.1
La definición gráfica de una derivada: conforme Δx se aproxima a cero al ir de *a*) a *c*), la aproximación por diferencias se va convirtiendo en una derivada.

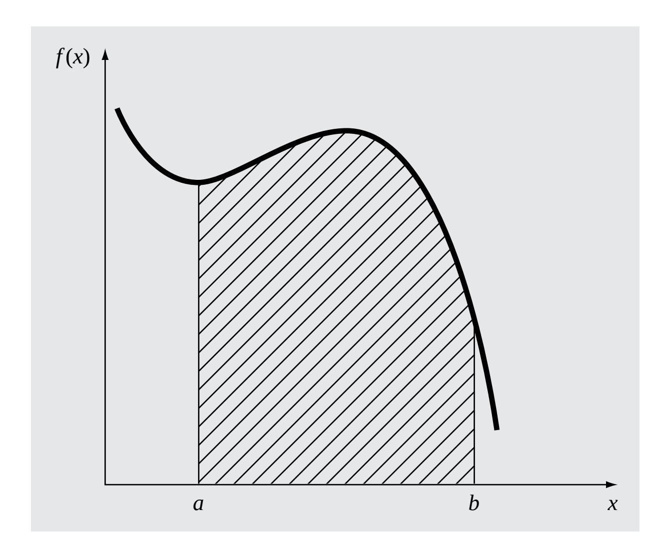


Figura PT6.2
Representación gráfica de la integral de $f(x)$ entre los límites $x = a$ y $x = b$. La integral es equivalente al área bajo la curva.

La *segunda derivada* representa la derivada de la primera derivada

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right)$$

Así, la segunda derivada expresa la rapidez de cambio de la pendiente. Con frecuencia se le llama la *curvatura*, porque un alto valor de la segunda derivada significa alta curvatura.

Finalmente, para funciones que dependen de más de una variable, se usan las *derivadas parciales*. Se puede pensar que éstas toman la derivada de la función en un punto manteniendo constantes todas las variables excepto una. Por ejemplo, dada una función f que depende tanto de x como de y, la derivada parcial de f con respecto a x en un punto arbitrario (x, y) se define como

$$\frac{\partial f}{\partial x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}$$

Del mismo modo, la derivada parcial de f con respecto a y se define como

$$\frac{\partial f}{\partial y} = \lim_{\Delta y \to 0} \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y}$$

Para adquirir una comprensión intuitiva de las derivadas parciales, reconozca que una función que depende de dos variables es una superficie más que una curva. Suponga que usted está escalando una montaña y tiene acceso a una función f que da la elevación como función de la longitud (el eje x, con orientación este-oeste) y la latitud (el eje y, con orientación norte-sur). Si usted se detiene en un punto determinado (x_0, y_0), la pendiente al este sería $\partial f(x_0, y_0)/\partial x$, y la pendiente al norte sería $\partial f(x_0, y_0)/\partial y$.

En cálculo, el proceso inverso de la diferenciación es la integración. De acuerdo con la definición del diccionario, *integrar* significa "juntar partes en un todo; unir; indicar la cantidad total ...". Matemáticamente, la integración se representa por

$$I = \int_a^b f(x)\,dx \tag{PT6.2}$$

que representa la integral de la función $f(x)$ con respecto a la variable independiente x, evaluada entre los límites $x = a$ y $x = b$. La función $f(x)$ en la ecuación (PT6.2) se llama *integrando*.

Como lo sugiere la definición del diccionario, el "significado" de la ecuación (PT6.2) es el *valor total*, o *sumatoria*, de $f(x)\,dx$ sobre el intervalo desde $x = a$ hasta $x = b$. De hecho, el símbolo ∫ es en realidad una letra *S* mayúscula estilizada, que intenta representar la estrecha relación entre integración y suma.

La figura PT6.2 representa una manifestación gráfica del concepto. Para funciones que están por encima del eje x, la integral, expresada por la ecuación (PT6.2) corresponde al área bajo la curva de $f(x)$ entre $x = a$ y b.[1]

Como se dijo antes, la "distinción" o "discriminación" de la diferenciación y el "juntar" de la integral son procesos estrechamente relacionados, de hecho, inversamente relacionados (figura PT6.3). Por ejemplo, si se tiene una función dada $y(t)$ que especifica la posición de un objeto en función del tiempo, la diferenciación proporciona un medio para determinar su velocidad (figura PT6.3*a*),

$$v(t) = \frac{d}{dt}y(t)$$

[1] Deberá observarse que el proceso representado por la ecuación (PT6.2) y la figura PT6.2 se conoce como *integración definida*. Hay otro tipo que se denomina integración indefinida, en la cual no se especifican los límites a y b. Como se analizará en la parte siete, la *integración indefinida* se ocupa de la determinación de una función, de la que se da su derivada.

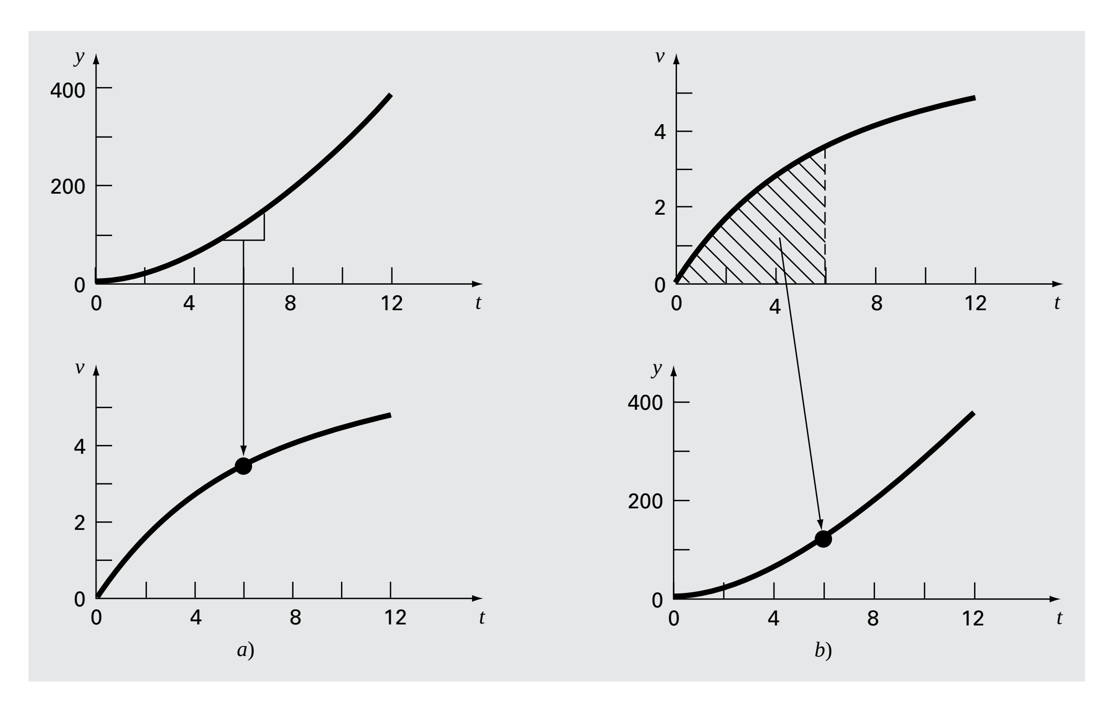


Figura PT6.3
El contraste entre *a*) diferenciación y *b*) integración.

De manera inversa, si se tiene la velocidad como una función del tiempo, la integración se utilizará para determinar su posición (figura PT6.3*b*),

$$y(t) = \int_0^t v(t)\, dt$$

Así, se dice de manera general que la evaluación de la integral

$$I = \int_a^b f(x)\, dx$$

es equivalente a resolver la ecuación diferencial

$$\frac{dy}{dx} = f(x)$$

para $y(b)$ dada la condición inicial $y(a) = 0$.

Debido a esa estrecha relación, optamos por dedicar esta parte del libro a ambos procesos. Entre otras cuestiones, esto ofrecerá la oportunidad de resaltar tanto sus similitudes como sus diferencias desde una perspectiva numérica. Además, nuestro análisis tendrá relevancia en las siguientes partes del libro, donde se estudiarán las ecuaciones diferenciales.

PT6.1.1 Métodos sin computadora para diferenciación e integración

La función que va a diferenciarse o integrarse estará, usualmente, en una de las siguientes tres formas:

1. Una función continua simple como un polinomio, una función exponencial o una función trigonométrica.
2. Una función continua complicada que es difícil o imposible de diferenciar o integrar directamente.
3. Una función tabulada donde los valores de x y $f(x)$ están dados como un conjunto discreto de puntos, lo cual es el caso cuando se tienen datos experimentales o de campo.

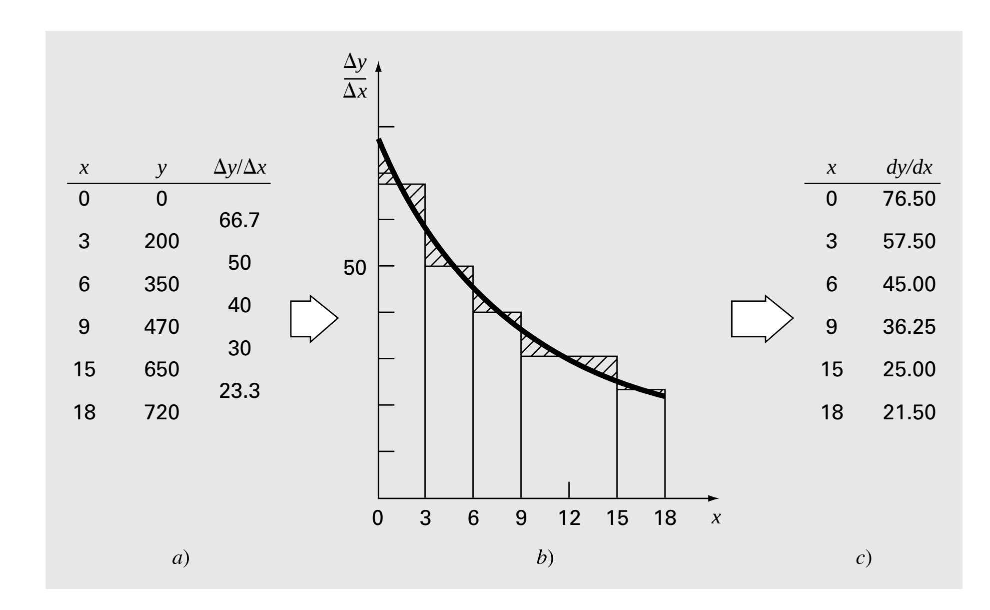


Figura PT6.4
Diferenciación por áreas iguales. *a*) Se usan las diferencias divididas centradas para estimar la derivada en cada intervalo entre los puntos asociados con los datos. *b*) Las estimaciones de la derivada se representan en forma de gráfica de barras. Se superpone una curva suave sobre esta gráfica para aproximar el área bajo la gráfica de barras. Esto se lleva a cabo al dibujar la curva de tal forma que áreas iguales positivas y negativas estén equilibradas. *c*) Entonces, es posible leer los valores de dy/dx de la curva suave.

En el primer caso, la derivada o la integral de una función simple se puede evaluar analíticamente usando el cálculo. En el segundo caso, las soluciones analíticas a menudo no son fáciles e incluso algunas veces son imposibles de obtener. En tales situaciones, así como en el tercer caso de datos discretos, se deberán emplear métodos aproximados.

Un método sin computadora para determinar las derivadas a partir de datos se conoce como *diferenciación gráfica por áreas iguales*. En este método los datos (x, y) se tabulan y, para cada intervalo, se emplea una diferencia dividida simple $\Delta y/\Delta x$ para estimar la pendiente. Después, esos valores se grafican como una curva escalonada contra x (figura PT6.4). Luego se dibuja una curva suave que trata de aproximar el área bajo la curva escalonada. Es decir, se dibuja de manera que las áreas negativas y positivas se equilibren visualmente. Entonces, las razones para valores dados de x pueden leerse en la curva.

De esta misma manera, se utilizaron procedimientos visualmente orientados para integrar datos tabulados y funciones complicadas, antes de la llegada de la computadora. Un procedimiento intuitivo simple consiste en graficar la función sobre una cuadrícula (figura PT6.5) y contar el número de cuadros que se aproximen al área. Este número multiplicado por el área de cada cuadro proporciona una burda estimación del área total bajo la curva. Dicha estimación se puede mejorar, a expensas de mayor trabajo, usando una cuadrícula más fina.

Otro procedimiento de sentido común es dividir el área en segmentos verticales, o barras, con una altura igual al valor de la función en el punto medio de cada barra (figura PT6.6). Después, el área de los rectángulos se calcula y se suma para estimar el área total. En este procedimiento se supone que el valor en el punto medio de la barra ofrece una aproximación válida de la altura promedio de la función en cada barra. Como en el método de la cuadrícula, es posible mejorar las estimaciones al usar más barras (y en consecuencia más delgadas) para aproximar la integral.

Aunque tales procedimientos tienen utilidad para estimaciones rápidas, existen técnicas numéricas alternativas con el mismo propósito. No es de sorprender entonces que los más simples de estos métodos sean similares, en esencia, a las técnicas sin computadora.

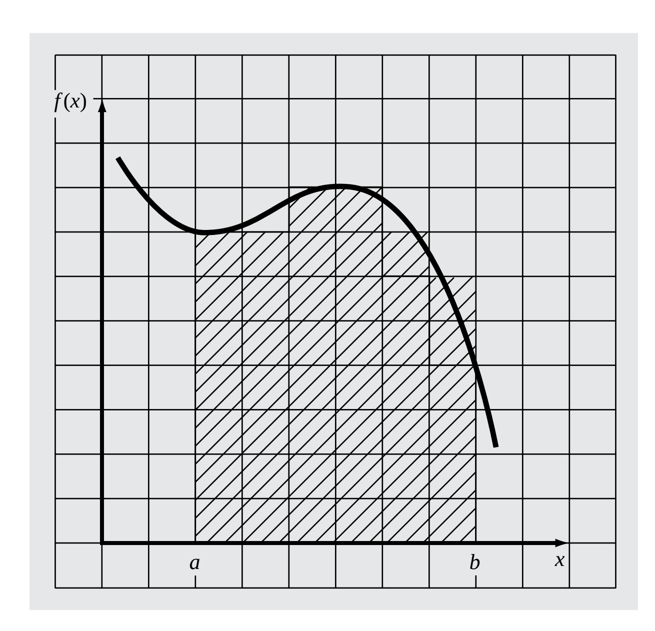


Figura PT6.5
Uso de una cuadrícula para aproximar una integral.

Para la diferenciación, las técnicas numéricas fundamentales utilizan divididas diferencias finitas para estimar las derivadas. Para datos con error, un procedimiento alternativo consiste en ajustar a los datos una curva suave con una técnica como la de regresión por mínimos cuadrados y luego derivar esta curva para obtener las estimaciones correspondientes.

De la misma forma, se dispone de integración numérica o de métodos de *cuadratura* para obtener integrales. Dichos métodos, que, de hecho, son más fáciles de implementar que el método de la cuadrícula, son similares en esencia al método por barras. Es decir, las alturas de la función se multiplican por el ancho de las barras y se suman para estimar la integral. Sin embargo, mediante una elección inteligente de los factores ponderantes, la estimación resultante se puede hacer más exacta que con el "método de barras" simple.

Como en el método de barras simple, las técnicas numéricas de integración y diferenciación utilizan datos de puntos discretos. Como cierta información ya está tabulada, naturalmente es compatible con muchos de los métodos numéricos. Aunque las funciones continuas no están originalmente en forma discreta, a menudo resulta sencillo emplear las ecuaciones dadas para generar una tabla de valores. Como se ilustra en la figura PT6.7, esta tabla puede, entonces, evaluarse con un método numérico.

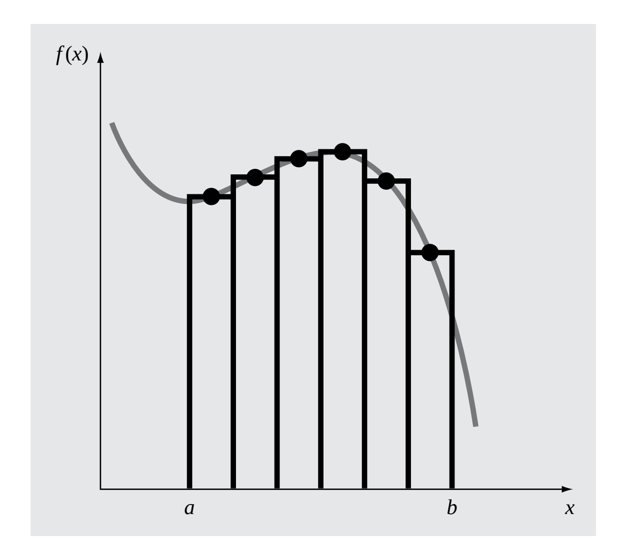


Figura PT6.6
El empleo de rectángulos, o barras, para aproximar la integral.

PT6.1.2 Diferenciación e integración numérica en ingeniería

La diferenciación e integración de una función tiene tantas aplicaciones en la ingeniería que usted tuvo que estudiar cálculo diferencial e integral en su primer año de estudios superiores. Se podrían dar muchos ejemplos específicos de tales aplicaciones en todos los campos de la ingeniería.

La diferenciación es algo común en ingeniería a causa de que mucho de nuestro trabajo implica analizar los cambios de las variables, tanto en el tiempo como en el espacio. De hecho, muchas de las leyes, y otras generalizaciones que aparecen constantemente en nuestro trabajo, se basan en las maneras predecibles donde el cambio se manifiesta en el mundo físico. Un ejemplo importante es la segunda ley de Newton, que no se expresa en términos de la posición de un objeto, sino más bien en el cambio de la posición con respecto al tiempo.

Además de este ejemplo que involucra el tiempo, numerosas leyes que gobiernan el comportamiento de las variables en el espacio se expresan en términos de derivadas. Entre las más comunes figuran las leyes que consideran potenciales o gradientes. Por ejemplo, la *ley de Fourier de la conducción de calor* cuantifica la observación de que el calor fluye desde regiones de mayor a menor temperatura. En el caso unidimensional, ésta se expresa en forma matemática como:

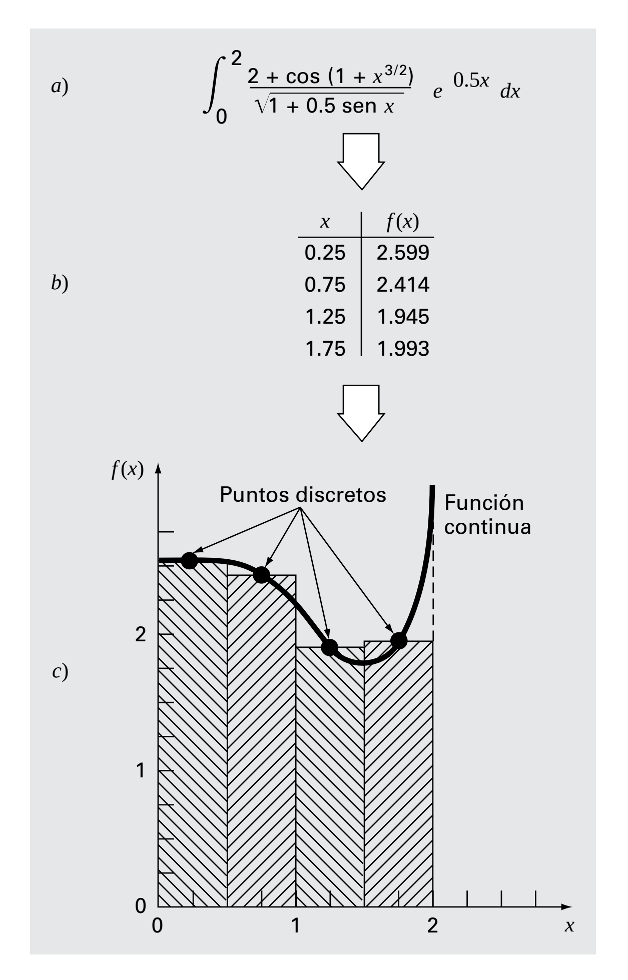


Figura PT6.7
Aplicación de un método de integración numérico: *a*) Una función continua complicada. *b*) Tabla de valores discretos de $f(x)$ generados a partir de la función. *c*) Uso de un método numérico (el método de barras) para estimar la integral basándose en puntos discretos. En una función tabulada, los datos ya están en esta forma *b*); por lo tanto, el paso *a*) no es necesario.

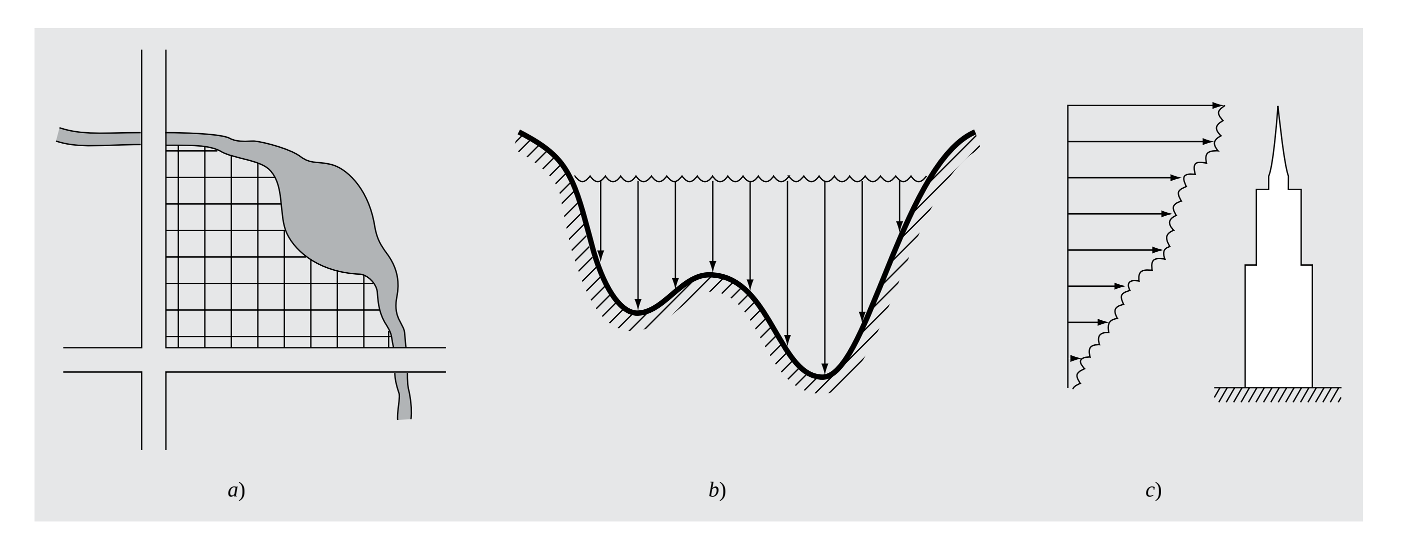


Figura PT6.8
Ejemplos de cómo se utiliza la integración para evaluar áreas en problemas de ingeniería. *a*) Un topógrafo podría necesitar saber el área de un campo limitado por una corriente zigzagueante y dos caminos. *b*) Un ingeniero en hidráulica tal vez requiera conocer el área de la sección transversal de un río. *c*) Un ingeniero en estructuras quizá necesite determinar la fuerza neta ejercida por un viento no uniforme que sopla contra un lado de un rascacielos.

$$\text{Flujo de calor} = -k' \frac{dT}{dx}$$

Así, la derivada proporciona una medida de la intensidad del cambio de temperatura, o *gradiente*, que ocasiona la transferencia de calor. Leyes similares proporcionan modelos prácticos en muchas áreas de la ingeniería, entre ellos se incluyen el modelado de dinámica de fluidos, la transferencia de masa, la cinética de las reacciones químicas y el flujo electromagnético. La habilidad para estimar de manera exacta las derivadas es una cualidad importante de nuestra capacidad para trabajar de manera eficiente en estas áreas.

Así como las estimaciones exactas de las derivadas son importantes en ingeniería, también el cálculo de integrales es igualmente valioso. Hay varios ejemplos relacionados directamente con la idea de la integral como el área bajo la curva. La figura PT6.8 ilustra algunos casos donde se usa la integración con este propósito.

Otras aplicaciones comunes relacionan la analogía entre integración y sumatoria. Por ejemplo, un problema común es determinar la media de funciones continuas. En la parte cinco se presentaron los conceptos de la media de n datos discretos [recuerde la ecuación PT5.1]:

$$\text{Media} = \frac{\sum_{i=1}^{n} y_i}{n} \tag{PT6.3}$$

donde y_i son las mediciones individuales. La determinación de la media para datos discretos se ilustra en la figura PT6.9*a*.

En contraste, suponga que y es una función continua de una variable independiente x, como se ilustra en la figura PT6.9*b*. En este caso existe un número infinito de valores entre a y b. Como la ecuación (PT6.3) se aplica para determinar la media de lecturas discretas, usted podría interesarse también en calcular la media o promedio de la función continua $y = f(x)$ en el intervalo de a a b. La integración se utiliza con este propósito, como lo especifica la fórmula

$$\text{Media} = \frac{\int_a^b f(x)\, dx}{b - a} \tag{PT6.4}$$

Esta fórmula tiene cientos de aplicaciones en ingeniería. Por ejemplo, sirve para calcular el centro de gravedad de objetos irregulares en ingeniería civil y mecánica, y para determinar la raíz media cuadrática de la corriente en ingeniería eléctrica.

Las integrales también se utilizan para evaluar la cantidad total de una variable física dada. La integral se puede evaluar sobre una línea, un área o un volumen. Por ejemplo, la masa total de una sustancia química contenida en un reactor está dada por el producto de la concentración de la sustancia química por el volumen del reactor, o

$$\text{Masa} = \text{concentración} \times \text{volumen}$$

donde la concentración tiene unidades de masa por volumen. Sin embargo, suponga que la concentración varía de un lugar a otro dentro del reactor. En este caso, es necesario sumar los productos de las concentraciones locales c_i por los correspondientes volúmenes elementales ΔV_i:

$$\text{Masa} = \sum_{i=1}^{n} c_i \, \Delta V_i$$

donde n es el número de volúmenes discretos. En el caso continuo, donde $c(x, y, z)$ es una función conocida y x, y, z son las variables independientes que designan la posición en coordenadas cartesianas, la integración se utiliza con el mismo propósito:

$$\text{Masa} = \iiint c(x, y, z) \, dx \, dy \, dz$$

o

$$\text{Masa} = \iiint_V c(V) \, dV$$

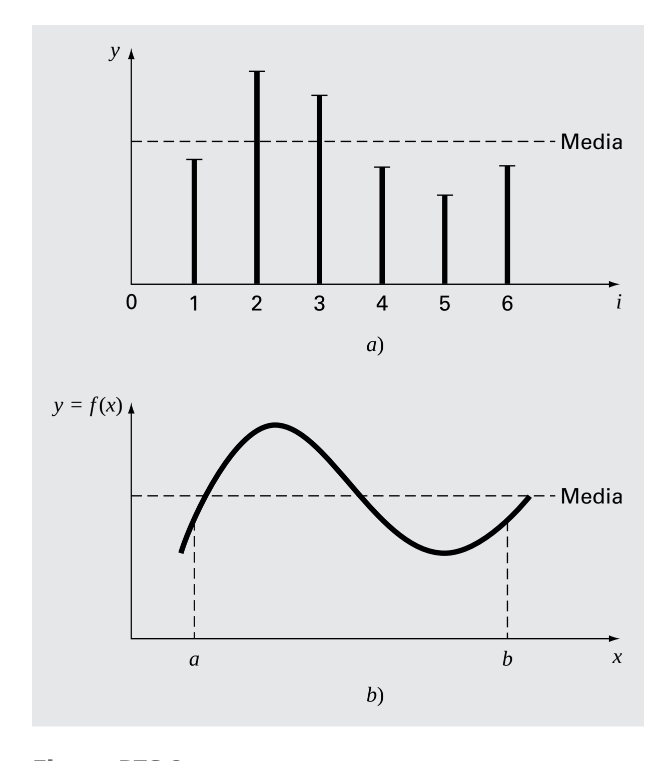


Figura PT6.9
Una ilustración de la media para datos *a*) discretos y *b*) continuos.

la cual se conoce como una *integral de volumen*. Observe la estrecha analogía entre suma e integración.

Casos similares podrán darse en otros campos de la ingeniería. Por ejemplo, la rapidez total de la transferencia de energía a través de un plano, donde el flujo (en calorías por centímetro cuadrado por segundo) es una función de la posición, está dada por

$$\text{Transferencia de calor} = \iint_A \text{flujo} \, dA$$

que se denomina una *integral de área*, donde A = área.

De manera similar, para el caso unidimensional, la masa total de una barra con densidad variable, y que tiene un área de sección transversal constante, está dada por

$$m = A \int_0^L \rho(x) \, dx$$

donde m = masa total (kg), L = longitud de la barra (m), $\rho(x)$ = densidad conocida (kg/m^3) como una función de la longitud x (m) y A = área de la sección transversal de la barra (m^2).

Por último, las integrales se utilizan para resolver ecuaciones diferenciales. Por ejemplo, suponga que la velocidad de una partícula es una función continua conocida del tiempo $v(t)$,

$$\frac{dy}{dt} = v(t)$$

La distancia total y recorrida por esta partícula en un tiempo t está dada por (figura PT6.3*b*)

$$y = \int_0^t v(t)\, dt \tag{PT6.5}$$

Éstas son sólo algunas de las diversas aplicaciones de la diferenciación y la integración que usted podría enfrentar regularmente durante el desarrollo de su profesión. Cuando las funciones sujetas a análisis son simples, usted preferirá evaluarlas analíticamente. Por ejemplo, en el problema del paracaidista en caída, determinamos la velocidad como función del tiempo [ecuación (1.10)]. Esta relación podría sustituirse en la ecuación (PT6.5), la cual se integra con facilidad, para determinar la distancia que cae el paracaidista en un periodo t. En un caso así, la integral es fácil de evaluar. Sin embargo, es difícil, o imposible, cuando la función es complicada, como sucede en el caso de ejemplos reales. Además, a menudo la función analizada se desconoce y se define, sólo por mediciones en puntos discretos. En ambos casos, usted debe tener la habilidad de obtener valores aproximados para las derivadas e integrales mediante técnicas numéricas. Varias de esas técnicas se analizarán en esta parte del libro.

PT6.2 ANTECEDENTES MATEMÁTICOS

En el nivel medio superior o durante su primer año en el nivel superior, se le dio una introducción al *cálculo diferencial e integral*. Ahí usted aprendió técnicas para obtener derivadas e integrales exactas o analíticas.

Cuando diferenciamos una función de manera analítica, generamos una segunda función que se utiliza para calcular la derivada de valores diferentes en la variable independiente. Existen reglas generales para este propósito. Por ejemplo, en el caso del monomio

$$y = x^n$$

se aplica la siguiente regla sencilla ($n \neq 0$):

$$\frac{dy}{dx} = nx^{n-1}$$

que es la expresión de la regla más general para

$$y = u^n$$

donde u = una función de x. En esta ecuación, la derivada se calcula usando la regla de la cadena

$$\frac{dy}{dx} = nu^{n-1}\frac{du}{dx}$$

Otras dos fórmulas se aplican a los productos o cocientes de funciones. Por ejemplo, si el producto de dos funciones de x(u y v) se representa como $y = uv$, entonces la derivada se calcula como

$$\frac{dy}{dx} = u\frac{dv}{dx} + v\frac{du}{dx}$$

Para la división, $y = u/v$, la derivada se calcula como

$$\frac{dy}{dx} = \frac{v\dfrac{du}{dx} - u\dfrac{dv}{dx}}{v^2}$$

Otras fórmulas útiles se resumen en la tabla PT6.1.

Existen fórmulas similares para la integración definida, donde se busca determinar una integral entre límites específicos, como en

Tabla PT6.1 Algunas derivadas de uso común.

$\frac{d}{dx}\operatorname{sen} x = \cos x$	$\frac{d}{dx}\cot x = -\csc^2 x$
$\frac{d}{dx}\cos x = -\operatorname{sen} x$	$\frac{d}{dx}\sec x = \sec x \tan x$
$\frac{d}{dx}\tan x = \sec^2 x$	$\frac{d}{dx}\csc x = -\csc x \cot x$
$\frac{d}{dx}\ln x = \frac{1}{x}$	$\frac{d}{dx}\log_a x = \frac{1}{x \ln a}$
$\frac{d}{dx}e^x = e^x$	$\frac{d}{dx}a^x = a^x \ln a$

$$I = \int_a^b f(x)\,dx \tag{PT6.6}$$

De acuerdo con el *teorema fundamental* del cálculo integral, la ecuación (PT6.6) se evalúa así

$$\int_a^b f(x)\,dx = F(x)\big|_a^b$$

donde $F(x)$ = integral de $f(x)$; es decir, cualquier función tal que $F'(x) = f(x)$. La nomenclatura del lado derecho corresponde a

$$F(x)\big|_a^b = F(b) - F(a) \tag{PT6.7}$$

Un ejemplo de una integral definida es

$$I = \int_0^{0.8} (0.2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5)\,dx \tag{PT6.8}$$

En este caso, la función es un simple polinomio que puede integrarse de manera analítica al calcular cada término de acuerdo con la regla

$$\int_a^b x^n\,dx = \frac{x^{n+1}}{n+1}\bigg|_a^b \tag{PT6.9}$$

donde n no puede ser igual a –1. Si se aplica esta regla a cada término de la ecuación (PT6.8) se obtiene

$$I = 0.2x + 12.5x^2 - \frac{200}{3}x^3 + 168.75x^4 - 180x^5 + \frac{400}{6}x^6\bigg|_0^{0.8} \tag{PT6.10}$$

la cual se evalúa de acuerdo con la ecuación (PT6.7) como $I = 1.6405333$. Este valor es igual al área bajo el polinomio original [ecuación (PT6.8)] entre $x = 0$ y 0.8.

La integración anterior depende del conocimiento de la regla expresada por la ecuación (PT6.9). Otras funciones siguen diferentes reglas. Estas "reglas" son sólo ejemplos de *antidiferenciación*; es decir, se busca encontrar $F(x)$ de tal forma que $F'(x) = f(x)$. En consecuencia, la integración analítica depende del conocimiento previo de la respuesta. Tal conocimiento se adquiere con entrenamiento y experiencia. Muchas de las reglas se resumen en manuales y tablas de integrales. Enlistamos algunas de las más comunes en la tabla PT6.2. Sin embargo, muchas funciones de importancia práctica son demasiado complicadas para estar contenidas en dicha tabla. Una razón por la que las técnicas en esta parte del libro son tan valiosas es porque ofrecen un medio para evaluar relaciones como la ecuación (PT6.8) sin conocimiento de las reglas.

PT6.3 ORIENTACIÓN

Antes de proceder con los métodos numéricos para la integración, podría ser de utilidad alguna orientación adicional. El siguiente material busca ofrecer una visión preliminar de los temas que se analizan en la parte seis. Además, formulamos algunos objetivos que ayudarán a centrar su atención cuando estudie la parte correspondiente.

PT6.3.1 Alcance y presentación preliminar

La figura PT6.10 proporciona una visión general de la parte seis. El *capítulo 21* se dedica al más común de los procedimientos para la integración numérica (las *fórmulas de Newton-Cotes*). Tales relaciones se basan en el reemplazo de una función complicada o para datos tabulados con un simple polinomio que es fácil de integrar. Tres de las fórmulas de Newton-Cotes más utilizadas se examinan con detalles: la *regla del trapecio*, la *regla de Simpson 1/3* y la *regla de Simpson 3/8*. Todas ellas se diseñaron para los casos donde los datos que van a integrarse están igualmente espaciados. Además,

Tabla PT6.2 Algunas integrales simples que se usan en la parte seis. En esta tabla las letras *a* y *b* son constantes y no deberán confundirse con los límites de integración analizados en el texto.

$$\int u\,dv = uv - \int v\,du$$

$$\int u^n\,du = \frac{u^{n+1}}{n+1} + C \qquad n \neq -1$$

$$\int a^{bx}\,dx = \frac{a^{bx}}{b \ln a} + C \qquad a > 0,\ a \neq 1$$

$$\int \frac{dx}{x} = \ln|x| + C \qquad x \neq 0$$

$$\int \operatorname{sen}(ax+b)\,dx = -\frac{1}{a}\cos(ax+b) + C$$

$$\int \cos(ax+b)\,dx = \frac{1}{a}\operatorname{sen}(ax+b) + C$$

$$\int \ln|x|\,dx = x\ln|x| - x + C$$

$$\int e^{ax}\,dx = \frac{e^{ax}}{a} + C$$

$$\int xe^{ax}\,dx = \frac{e^{ax}}{a^2}(ax-1) + C$$

$$\int \frac{dx}{a+bx^2} = \frac{1}{\sqrt{ab}}\tan^{-1}\frac{\sqrt{ab}}{a}x + C$$

incluimos un análisis de la integración numérica de datos irregularmente espaciados. Este tema es muy importante, ya que muchos de los problemas del mundo real tienen que ver con datos espaciados de esta manera.

Todo el material anterior se relaciona con la integración cerrada; es decir, cuando se conocen los valores de la función en los límites de integración. Al final del capítulo 21 presentamos *fórmulas de integración abierta*; es decir, donde los límites de integración se extienden más allá del rango de los datos conocidos. Aunque estas fórmulas no se usan comúnmente para la integración definida, las fórmulas de integración abierta se presentan aquí debido a que se utilizan bastante en la parte siete para la solución de ecuaciones diferenciales ordinarias.

Las formulaciones vistas en el capítulo 21 pueden utilizarse para analizar tanto los datos tabulados como las ecuaciones. El *capítulo 22* se ocupa de tres técnicas que están expresamente diseñadas para integrar ecuaciones y funciones: la *integración de Romberg*, la *cuadratura adaptativa*, y la *cuadratura de Gauss*. Se proporcionan algoritmos computacionales para estos métodos. Además, se analizan métodos para evaluar *integrales impropias*.

En el *capítulo 23* se presenta información adicional sobre *diferenciación numérica* para complementar el material introductorio del capítulo 4. Los temas comprenden las fórmulas por diferencias finitas de alta precisión, la extrapolación de Richardson y la diferenciación de datos irregularmente espaciados. Se analiza el efecto de los errores tanto en la diferenciación numérica como en la integración. Por último, se concluye el capítulo con una descripción de la aplicación de diferentes paquetes de software para integración y diferenciación.

El *capítulo 24* muestra cómo se aplican los métodos a la solución de problemas. De la misma manera que en las otras partes del libro, los problemas se toman de los campos de la ingeniería.

Una sección de repaso, o *epílogo*, se presenta al final de la parte seis. Esta revisión comprende un análisis de las ventajas y las desventajas que son relevantes para la implementación en la práctica de la ingeniería. Además, se resumen fórmulas importantes. Por último, se presenta una breve revisión de los métodos avanzados y las referencias alternativas que facilitarán sus estudios posteriores sobre la diferenciación y la integración numérica.

PT6.3.2 Metas y objetivos

Objetivos de estudio Después de terminar la parte seis, usted será capaz de resolver muchos problemas de integración y diferenciación numérica y darse cuenta del valor de su aplicación en la solución de problemas en ingeniería. También deberá esforzarse por dominar diferentes técnicas y evaluar su confiabilidad. Usted deberá comprender las ventajas y las desventajas al seleccionar el "mejor" método (o métodos), para cualquier problema específico. Además de estos objetivos generales, deberá asimilar y dominar los conceptos específicos que se presentan en la tabla PT6.3.

Objetivos de cómputo Se le han proporcionado software y algoritmos computacionales simples para implementar las técnicas analizadas en la parte seis. Todo esto tiene utilidad como herramienta de aprendizaje.

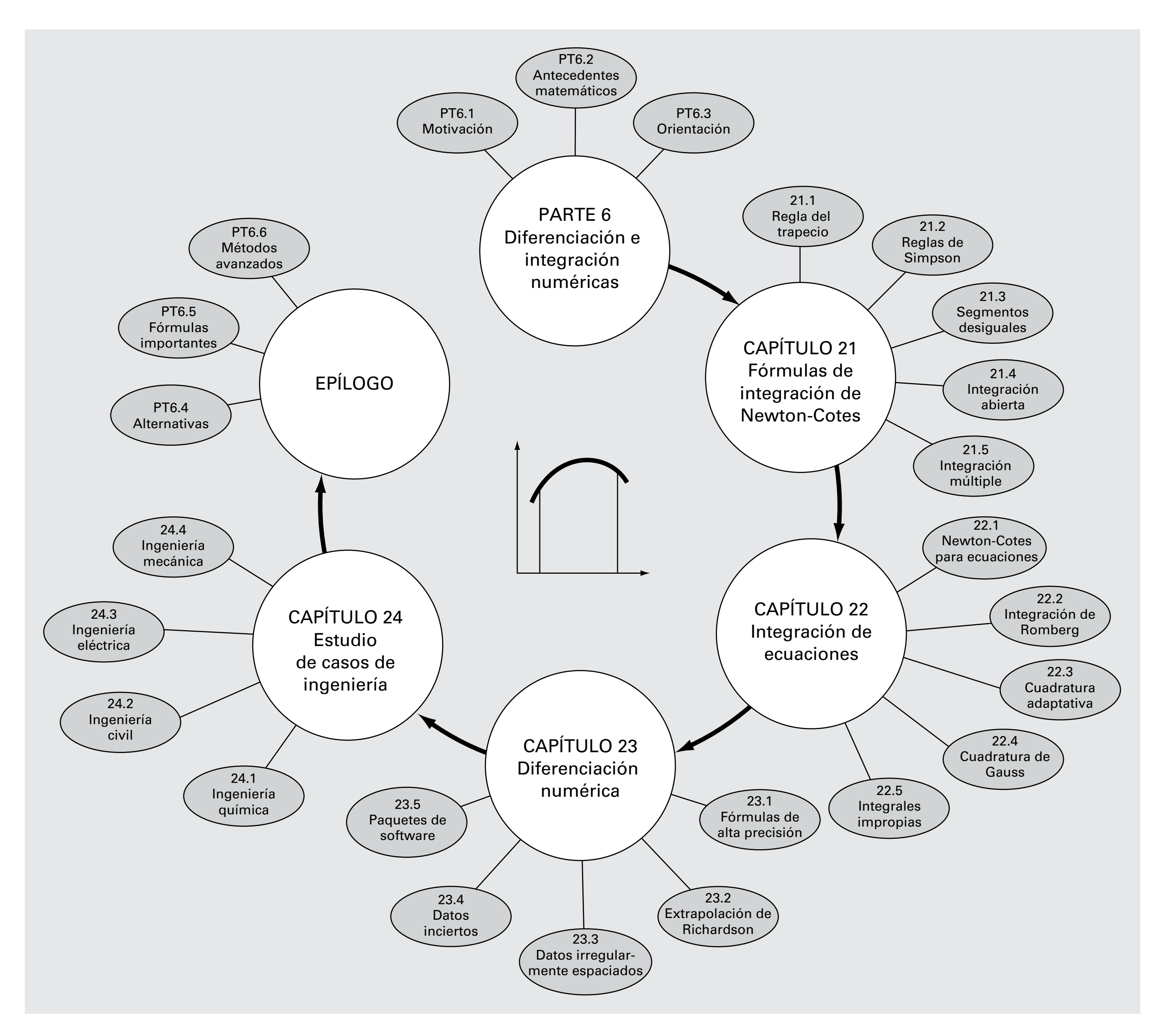


Figura PT6.10
Esquema de la organización del material en la parte seis: Diferenciación e integración numéricas.

Además, se proporcionan algoritmos para la mayoría de los otros métodos de la parte seis. Esta información le permitirá ampliar su software al incluir técnicas más allá de la regla del trapecio. Por ejemplo, quizá encuentre útil, desde un punto de vista profesional, tener software para implementar la integración y la diferenciación numéricas para datos irregularmente espaciados. También podrá desarrollar su propio software para las reglas de Simpson, la integración de Romberg, la integración adaptativa, y la cuadratura de Gauss, que usualmente son más eficientes y exactos que la regla del trapecio.

Tabla PT6.3 Objetivos específicos de estudio de la parte seis.

1. Entender la obtención de las fórmulas de Newton-Cotes; saber cómo obtener la regla del trapecio y cómo obtener las reglas de Simpson; reconocer que las reglas del trapecio y las de Simpson 1/3 y 3/8 representan las áreas bajo los polinomios de primero, segundo y tercer grados, respectivamente.
2. Conocer las fórmulas y las ecuaciones de error para *a*) la regla del trapecio, *b*) la regla del trapecio de aplicación múltiple, *c*) la regla de Simpson 1/3, *d*) la regla de Simpson 3/8 y *e*) la regla de Simpson de aplicación múltiple. Ser capaz de elegir la “mejor” de estas fórmulas para cualquier contexto de un problema específico.
3. Comprender que la regla de Simpson 1/3 tiene una exactitud de cuarto orden, aun cuando se base en sólo tres puntos; darse cuenta de que todas las fórmulas de Newton-Cotes de segmentos pares y puntos impares tienen exactitud mejorada similar.
4. Saber cómo evaluar la integral y la derivada de datos desigualmente espaciados.
5. Reconocer la diferencia entre las fórmulas de integración abierta y cerrada.
6. Comprender cómo evaluar numéricamente integrales múltiples.
7. Entender la base teórica de la extrapolación de Richardson, y cómo se aplica en el algoritmo de integración Romberg y en diferenciación numérica.
8. Distinguir la diferencia fundamental entre las fórmulas de Newton-Cotes y de cuadratura de Gauss.
9. Explicar por qué la integración de Romberg y la cuadratura de Gauss tienen utilidad cuando se integran ecuaciones (a diferencia de datos tabulares o discretos).
10. Saber cómo se emplean las fórmulas de integración abierta para evaluar integrales impropias.
11. Entender la aplicación de fórmulas de diferenciación numérica de alta precisión.
12. Saber cómo diferenciar datos desigualmente espaciados.
13. Reconocer los diferentes efectos del error en los datos para los procesos de integración y diferenciación numéricas.

Por último, una de las metas más importantes deberá ser dominar varios de los paquetes de software de uso general que están disponibles. En particular, usted deberá habituarse a usar estas herramientas para implementar los métodos numéricos en la solución de problemas de ingeniería.

CAPÍTULO 21

Fórmulas de integración de Newton-Cotes

Las *fórmulas* de *Newton-Cotes* son los tipos de integración numérica más comunes. Se basan en la estrategia de reemplazar una función complicada o datos tabulados por un polinomio de aproximación que es fácil de integrar:

$$I = \int_a^b f(x)\, dx \cong \int_a^b f_n(x)\, dx \tag{21.1}$$

donde $f_n(x)$ = un polinomio de la forma

$$f_n(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + a_n x^n$$

donde n es el grado del polinomio. Por ejemplo, en la figura 21.1*a*, se utiliza un polinomio de primer grado (una línea recta) como una aproximación. En la figura 21.1*b*, se emplea una parábola con el mismo propósito.

La integral también se puede aproximar usando un conjunto de polinomios aplicados por pedazos a la función o datos, sobre segmentos de longitud constante. Por ejemplo, en la figura 21.2, se usan tres segmentos de línea recta para aproximar la integral. Aunque pueden utilizarse polinomios de grado superior con los mismos propósitos. Con este antecedente, reconocemos que el "método de barras" de la figura PT6.6 emplea un conjunto de polinomios de grado cero (es decir, constantes) para aproximar la integral.

Existen formas cerradas y abiertas de las fórmulas de Newton-Cotes. Las *formas cerradas* son aquellas donde se conocen los datos al inicio y al final de los límites de integración (figura 21.3*a*). Las *formas abiertas* tienen límites de integración que se extienden más allá del intervalo de los datos (figura 21.3*b*). En este sentido, son similares a la extrapolación que se analizó en la sección 18.5. Por

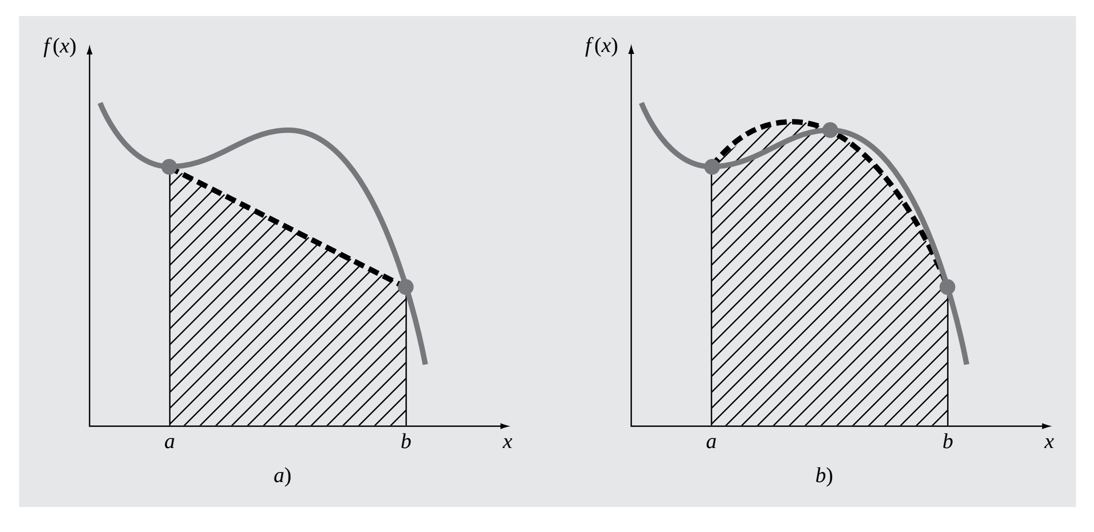


Figura 21.1
La aproximación de una integral mediante el área bajo *a*) una sola línea recta y *b*) una parábola.

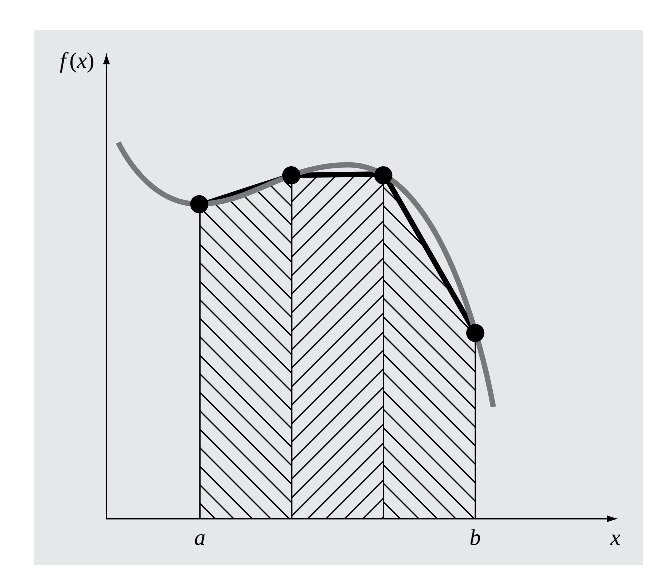


Figura 21.2
La aproximación de una integral mediante el área bajo tres segmentos de línea recta.

lo general, las formas abiertas de Newton-Cotes no se usan para integración definida. Sin embargo, se utilizan para evaluar integrales impropias y para obtener la solución de ecuaciones diferenciales ordinarias. Este capítulo enfatiza las formas cerradas. No obstante, al final del mismo se presenta brevemente una introducción a las fórmulas abiertas de Newton-Cotes.

21.1 LA REGLA DEL TRAPECIO

La *regla del trapecio* es la primera de las fórmulas cerradas de integración de Newton-Cotes. Corresponde al caso donde el polinomio de la ecuación (21.1) es de primer grado:

$$I = \int_a^b f(x)\,dx \cong \int_a^b f_1(x)\,dx$$

Recuerde del capítulo 18 que una línea recta se puede representar como [véase ecuación (18.2)]

$$f_1(x) = f(a) + \frac{f(b) - f(a)}{b - a}(x - a) \tag{21.2}$$

El área bajo esta línea recta es una aproximación de la integral de $f(x)$ entre los límites a y b:

$$I = \int_a^b \left[f(a) + \frac{f(b) - f(a)}{b - a}(x - a) \right] dx$$

El resultado de la integración (véase el cuadro 21.1 para detalles) es

$$I = (b - a)\frac{f(a) + f(b)}{2} \tag{21.3}$$

que se denomina *regla del trapecio.*

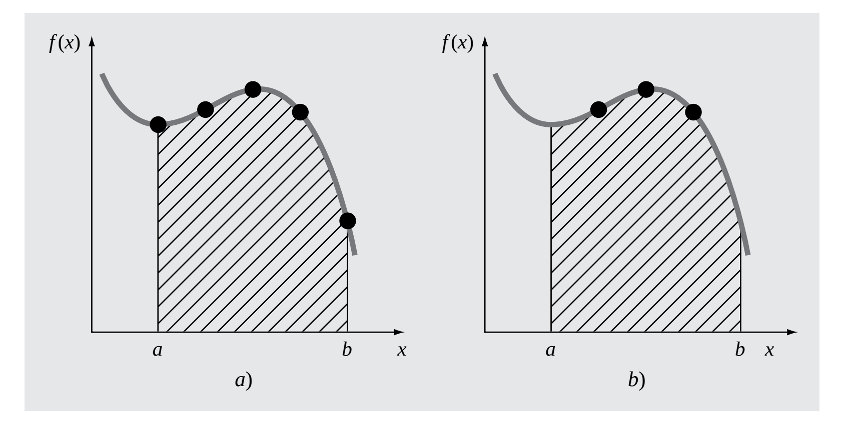


Figura 21.3
La diferencia entre las fórmulas de integración *a*) cerradas y *b*) abiertas.

Cuadro 21.1 Obtención de la regla del trapecio

Antes de la integración, la ecuación (21.2) se puede expresar como

$$f_1(x) = \frac{f(b)-f(a)}{b-a}x + f(a) - \frac{af(b)-af(a)}{b-a}$$

Agrupando los últimos dos términos:

$$f_1(x) = \frac{f(b)-f(a)}{b-a}x + \frac{bf(a)-af(a)-af(b)+af(a)}{b-a}$$

o

$$f_1(x) = \frac{f(b)-f(a)}{b-a}x + \frac{bf(a)-af(b)}{b-a}$$

la cual puede integrarse entre $x = a$ y $x = b$ para obtener:

$$I = \frac{f(b)-f(a)}{b-a}\frac{x^2}{2} + \frac{bf(a)-af(b)}{b-a}x\Big|_a^b$$

Este resultado se evalúa para dar:

$$I = \frac{f(b)-f(a)}{b-a}\frac{(b^2-a^2)}{2} + \frac{bf(a)-af(b)}{b-a}(b-a)$$

Ahora, como $b^2 - a^2 = (b - a)(b + a)$,

$$I = [f(b)-f(a)]\frac{b+a}{2} + bf(a) - af(b)$$

Multiplicando y agrupando términos se tiene:

$$I = (b-a)\frac{f(a)+f(b)}{2}$$

que es la fórmula para la regla del trapecio.

Geométricamente, la regla del trapecio es equivalente a aproximar el área del trapecio bajo la línea recta que une $f(a)$ y $f(b)$ en la figura 21.4. Recuerde que la fórmula para calcular el área de un trapecio es la altura por el promedio de las bases (figura 21.5*a*). En nuestro caso, el concepto es el mismo, pero el trapecio está sobre su lado (figura 21.5*b*). Por lo tanto, la integral aproximada se representa como

$$I \cong \text{ancho} \times \text{altura promedio} \tag{21.4}$$

o

$$I \cong (b-a) \times \text{altura promedio} \tag{21.5}$$

donde, para la regla del trapecio, la altura promedio es el promedio de los valores de la función en los puntos extremos, o $[f(a) + f(b)]/2$.

Todas las fórmulas cerradas de Newton-Cotes se expresan en la forma general de la ecuación (21.5). De hecho, sólo difieren respecto a la formulación de la altura promedio.

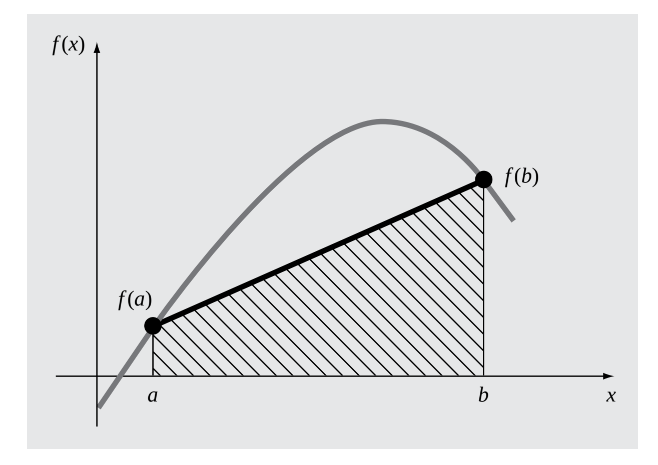


Figura 21.4
Representación gráfica de la regla del trapecio.

21.1.1 Error de la regla del trapecio

Cuando empleamos la integral bajo un segmento de línea recta para aproximar la integral bajo una curva, obviamente se tiene un error que puede ser importante (figura 21.6). Una estimación al error de truncamiento local para una sola aplicación de la regla del trapecio es (cuadro 21.2).

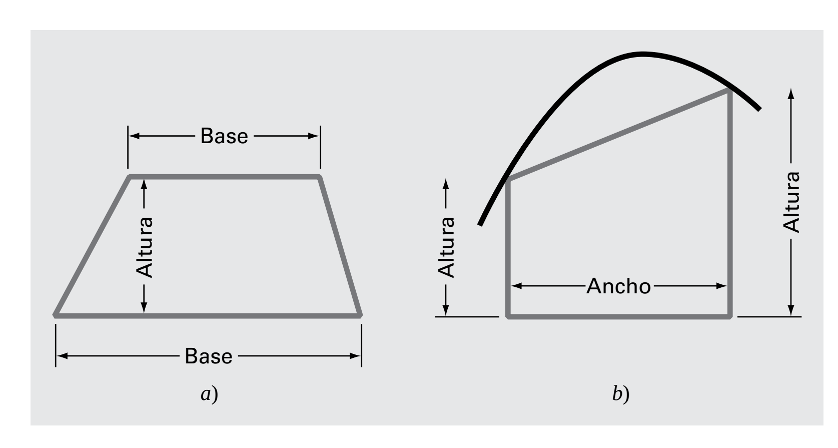


Figura 21.5
a) La fórmula para calcular el área de un trapecio: altura por el promedio de las bases. *b*) Para la regla del trapecio, el concepto es el mismo pero ahora el trapecio está sobre su lado.

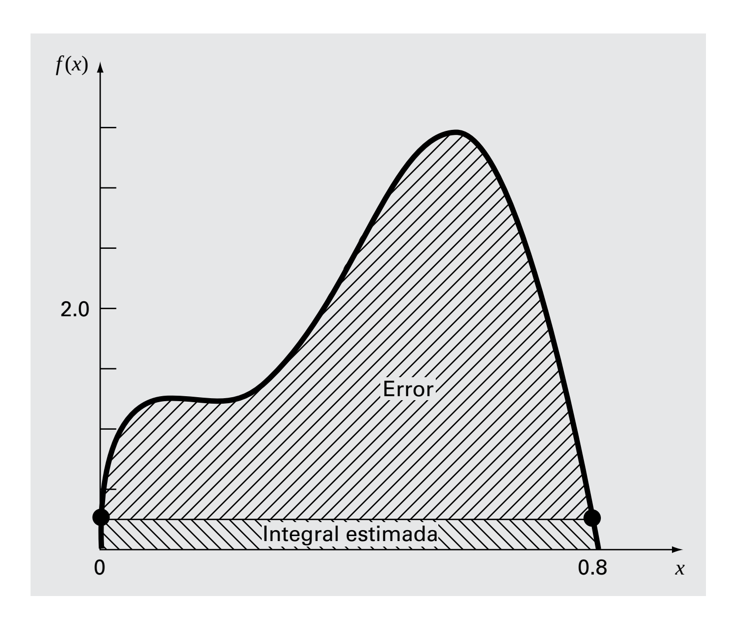


Figura 21.6
Representación gráfica del empleo de una sola aplicación de la regla del trapecio para aproximar la integral de $f(x) = 0.2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$ de $x = 0$ a 0.8.

Cuadro 21.2 Obtención y error estimado de la regla del trapecio

Una manera alternativa para obtener la regla del trapecio consiste en integrar el polinomio de interpolación hacia adelante de Newton-Gregory. Recuerde que para la versión de primer grado con el término del error, la integral será (cuadro 18.2)

$$I = \int_a^b \left[f(a) + \Delta f(a)\alpha + \frac{f''(\xi)}{2}\alpha(\alpha-1)h^2 \right] dx \qquad \text{(C21.2.1)}$$

Para simplificar el análisis, considere que si $\alpha = (x - a)/h$, entonces

$$dx = h\, d\alpha$$

Debido a que $h = b - a$ (para un segmento de la regla del trapecio), los límites de integración a y b corresponden a 0 y 1, respectivamente. Por lo tanto, la ecuación (C21.2.1) se expresará como

$$I = h\int_0^1 \left[f(a) + \Delta f(a)\alpha + \frac{f''(\xi)}{2}\alpha(\alpha-1)h^2 \right] d\alpha$$

Si se supone que para una h pequeña, el término $f''(\xi)$ es aproximadamente constante, entonces el resultado de la integración es:

$$I = h\left[\alpha\, f(a) + \frac{\alpha^2}{2}\Delta f(a) + \left(\frac{\alpha^3}{6} - \frac{\alpha^2}{4}\right) f''(\xi)h^2 \right]_0^1$$

y evaluado como

$$I = h\left[f(a) + \frac{\Delta f(a)}{2} \right] - \frac{1}{12} f''(\xi)h^3$$

Como $\Delta f(a) = f(b) - f(a)$, el resultado puede escribirse como

$$I = \underbrace{h\,\frac{f(a)+f(b)}{2}}_{\text{Regla del trapecio}} \underbrace{- \frac{1}{12} f''(\xi)h^3}_{\text{Error de truncamiento}}$$

Así, el primer término es la regla del trapecio y el segundo es una aproximación para el error.

$$E_t = -\frac{1}{12} f''(\xi)(b-a)^3 \qquad (21.6)$$

donde ξ está en algún lugar en el intervalo de a a b. La ecuación (21.6) indica que si la función sujeta a integración es lineal, la regla del trapecio será exacta. De otra manera, para funciones con derivadas de segundo orden y de orden superior (es decir, con curvatura), puede ocurrir algún error.

EJEMPLO 21.1 Aplicación simple de la regla del trapecio

Planteamiento del problema Con la ecuación (21.3) integre numéricamente

$$f(x) = 0.2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$$

desde $a = 0$ hasta $b = 0.8$. Recuerde de la sección PT6.2 que el valor exacto de la integral se puede determinar en forma analítica y es 1.640533.

Solución Al evaluar la función

$f(0) = 0.2$

$f(0.8) = 0.232$

sustituyendo en la ecuación (21.3) se tiene

$$I \cong 0.8 \frac{0.2 + 0.232}{2} = 0.1728$$

la cual representa un error de

$$E_t = 1.640533 - 0.1728 = 1.467733$$

que corresponde a un error relativo porcentual de $\varepsilon_t = 89.5\%$. La razón de este error tan grande es evidente en la gráfica de la figura 21.6. Observe que el área bajo la línea recta no toma en cuenta una porción significativa de la integral que está por encima de la línea.

En situaciones reales, tal vez no conozcamos previamente el valor verdadero. Por lo tanto, se requiere una estimación del error aproximado. Para obtener dicha estimación se calcula la segunda derivada de la función en el intervalo, derivando dos veces la función original:

$$f''(x) = -400 + 4\,050x - 10\,800x^2 + 8\,000x^3$$

El valor promedio de la segunda derivada se calcula mediante la ecuación (PT6.4):

$$\bar{f}''(x) = \frac{\int_0^{0.8} (-400 + 4\,050x - 10\,800x^2 + 8\,000x^3)\,dx}{0.8 - 0} = -60$$

que se sustituye en la ecuación (21.6) y el resultado es

$$E_a = -\frac{1}{12}(-60)(0.8)^3 = 2.56$$

que es del mismo orden de magnitud y signo que el error verdadero. Sin embargo, de hecho, existe una discrepancia, ya que en un intervalo de este tamaño, el promedio de la segunda derivada no es necesariamente una aproximación exacta de $f''(\xi)$. Así, indicamos que el error es aproximado mediante la notación E_a, y no exacto usando E_t.

21.1.2 La regla del trapecio de aplicación múltiple

Una forma de mejorar la precisión de la regla del trapecio consiste en dividir el intervalo de integración de a a b en varios segmentos, y aplicar el método a cada uno de ellos (figura 21.7). Las áreas de los segmentos se suman después para obtener la integral en todo el intervalo. Las ecuaciones resultantes se llaman *fórmulas de integración, de aplicación múltiple* o *compuestas*.

La figura 21.8 muestra el formato general y la nomenclatura que usaremos para obtener integrales de aplicación múltiple. Hay $n + 1$ puntos igualmente espaciados ($x_0, x_1, x_2,..., x_n$). En consecuencia, existen n segmentos del mismo ancho:

$$h = \frac{b - a}{n} \qquad (21.7)$$

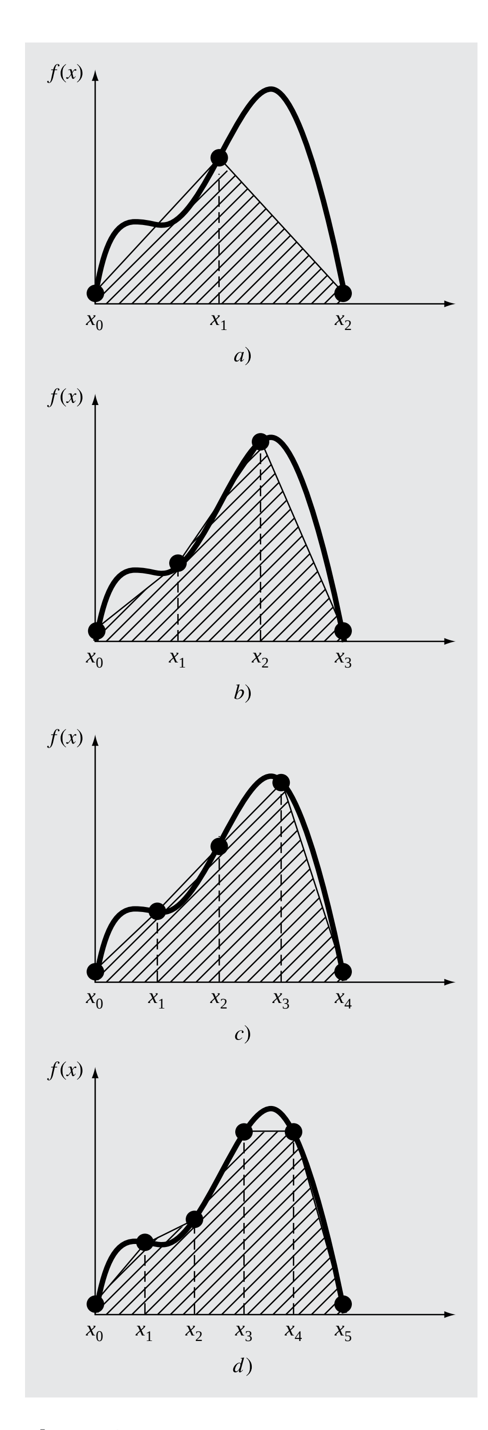


Figura 21.7
Ilustración de la regla del trapecio de aplicación múltiple. *a*) Dos segmentos, *b*) tres segmentos, *c*) cuatro segmentos y *d*) cinco segmentos.

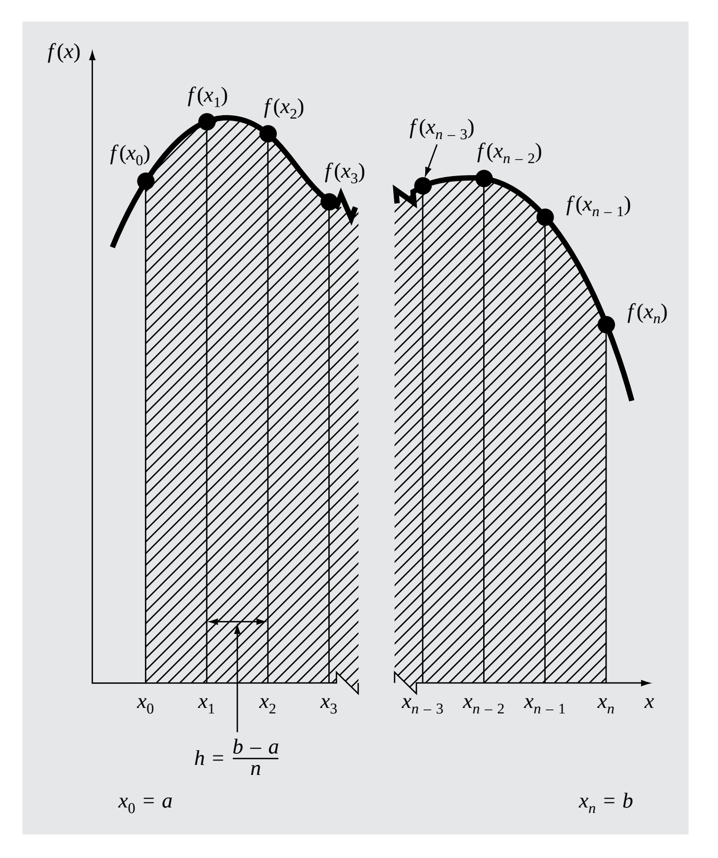


Figura 21.8
Formato general y nomenclatura para integrales de aplicación múltiple.

Si a y b se designan como x_0 y x_n, respectivamente, la integral completa se representará como

$$I = \int_{x_0}^{x_1} f(x)\,dx + \int_{x_1}^{x_2} f(x)\,dx + \cdots + \int_{x_{n-1}}^{x_n} f(x)\,dx$$

Sustituyendo la regla del trapecio en cada integral se obtiene

$$I = h\frac{f(x_0)+f(x_1)}{2} + h\frac{f(x_1)+f(x_2)}{2} + \cdots + h\frac{f(x_{n-1})+f(x_n)}{2} \tag{21.8}$$

o, agrupando términos,

$$I = \frac{h}{2}\left[f(x_0) + 2\sum_{i=1}^{n-1} f(x_i) + f(x_n) \right] \tag{21.9}$$

o, usando la ecuación (21.7) para expresar la ecuación (21.9) en la forma general de la ecuación (21.5),

$$I = \underbrace{(b-a)}_{\text{Ancho}}\underbrace{\frac{f(x_0) + 2\sum_{i=1}^{n-1} f(x_i) + f(x_n)}{2n}}_{\text{Altura promedio}} \tag{21.10}$$

Como la sumatoria de los coeficientes de $f(x)$ en el numerador dividido entre $2n$ es igual a 1, la altura promedio representa un promedio ponderado de los valores de la función. De acuerdo con la ecuación (21.10), a los puntos interiores se les da el doble de peso que a los dos puntos extremos $f(x_0)$ y $f(x_n)$.

Se tiene un error con la regla del trapecio de aplicación múltiple al sumar los errores individuales de cada segmento, así

$$E_t = -\frac{(b-a)^3}{12n^3}\sum_{i=1}^{n} f''(\xi_i) \tag{21.11}$$

donde $f''(\xi_i)$ es la segunda derivada en un punto ξ_i, localizado en el segmento i. Este resultado se simplifica al estimar la media o valor promedio de la segunda derivada en todo el intervalo como [ecuación (PT6.3)]

$$\bar{f}'' \cong \frac{\sum_{i=1}^{n} f''(\xi_i)}{n} \tag{21.12}$$

Por lo tanto, $\Sigma f''(\xi_i) \cong n\bar{f}''$ y la ecuación (21.11) se reescribe como

$$E_a = -\frac{(b-a)^3}{12n^2}\bar{f}'' \tag{21.13}$$

Así, si se duplica el número de segmentos, el error de truncamiento se divide entre cuatro. Observe que la ecuación (21.13) es un error aproximado debido a la naturaleza aproximada de la ecuación (21.12).

EJEMPLO 21.2 Regla del trapecio de aplicación múltiple

Planteamiento del problema Use la regla del trapecio con dos segmentos para estimar la integral de

$$f(x) = 0.2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$$

desde $a = 0$ hasta $b = 0.8$. Emplee la ecuación (21.13) para estimar el error. Recuerde que el valor correcto para la integral es 1.640533.

Solución $n = 2(h = 0.4)$:

$$f(0) = 0.2 \qquad f(0.4) = 2.456 \qquad f(0.8) = 0.232$$

$$I = 0.8\frac{0.2 + 2(2.456) + 0.232}{4} = 1.0688$$

$$E_t = 1.640533 - 1.0688 = 0.57173 \qquad \varepsilon_t = 34.9\%$$

$$E_a = -\frac{0.8^3}{12(2)^2}(-60) = 0.64$$

donde –60 es el promedio de la segunda derivada, determinada anteriormente en el ejemplo 21.1.

Los resultados del ejemplo anterior, junto con aplicaciones de la regla del trapecio con tres a diez segmentos, se resumen en la tabla 21.1. Observe cómo el error disminuye conforme aumenta el número de segmentos. Sin embargo, advierta también que la razón de disminución es gradual, a causa de que el error está relacionado inversamente con el cuadrado de n [ecuación (21.13)]. Por lo tanto, al

Tabla 21.1 Resultados de la regla del trapecio de aplicación múltiple para estimar la integral de $f(x) = 0.2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$ de $x = 0$ a 0.8. El valor exacto es 1.640533.

n	*h*	*I*	ε_t (%)
2	0.4	1.0688	34.9
3	0.2667	1.3695	16.5
4	0.2	1.4848	9.5
5	0.16	1.5399	6.1
6	0.1333	1.5703	4.3
7	0.1143	1.5887	3.2
8	0.1	1.6008	2.4
9	0.0889	1.6091	1.9
10	0.08	1.6150	1.6

duplicar el número de segmentos, el error se divide entre cuatro. En las siguientes secciones desarrollaremos fórmulas de grado superior que son más exactas y que convergen más rápido hacia la verdadera integral conforme los segmentos aumentan. Sin embargo, antes de investigar tales fórmulas, analizaremos algoritmos computacionales para implementar la regla del trapecio.

21.1.3 Algoritmos computacionales para la regla del trapecio

En la figura 21.9 se dan dos algoritmos simples para la regla del trapecio. El primero (figura 21.9*a*) es para la versión de un solo segmento. El segundo (figura 21.9*b*) es para la versión de múltiples segmentos con un ancho de segmento constante. Observe que ambos están diseñados para datos que se hallan en forma tabular. Un programa general deberá tener la capacidad de evaluar también funciones o ecuaciones conocidas. En el siguiente capítulo ilustraremos cómo se manipulan las funciones.

***a*) Un solo segmento**

```
FUNCTION Trap (h, f0, f1)
  Trap = h * (f0 + f1)/2
END Trap
```

***b*) Segmentos múltiples**

```
FUNCTION Trapm (h, n, f)
  sum = f0
  DOFOR i = 1, n – 1
    sum = sum + 2 * fi
  END DO
  sum = sum + fn
  Trapm = h * sum/2
END Trapm
```

Figura 21.9
Algoritmos para la regla del trapecio *a*) de un solo segmento y *b*) de múltiples segmentos.

EJEMPLO 21.3 Evaluación de integrales con la computadora

Planteamiento del problema Con software basado en la figura 21.9*b* resuelva un problema relacionado con el ya conocido: paracaidista en caída. Como usted recordará del ejemplo 1.1, la velocidad del paracaidista está dada con la siguiente función en términos del tiempo:

$$v(t) = \frac{gm}{c}(1 - e^{-(c/m)t}) \tag{E21.3.1}$$

donde v = velocidad (m/s), g = constante gravitacional de 9.8 m/s², m = masa del paracaidista igual a 68.1 kg y c = coeficiente de arrastre de 12.5 kg/s. El modelo predice la velocidad del paracaidista como una función del tiempo, de la manera en que se describió en el ejemplo 1.1.

Suponga que desea saber qué tan lejos ha caído el paracaidista después de cierto tiempo t. Tal distancia está determinada por [ecuación (PT6.5)]

$$d = \int_0^t v(t)\,dt$$

donde d es la distancia en metros. Sustituyendo en la ecuación (E21.3.1),

$$d = \frac{gm}{c}\int_0^t (1 - e^{-(c/m)t})\,dt$$

Use su propio software, para determinar esta integral mediante la regla del trapecio de aplicación múltiple con diferentes números de segmentos. Observe que realizando la integración en forma analítica y sustituyendo los valores de los parámetros conocidos se obtiene un valor exacto de d = 289.43515 m.

Solución En el caso en que n = 10 se obtiene una integral calculada de 288.7491. Así, hemos obtenido la integral con tres cifras significativas de exactitud. Los resultados con otros números de segmentos son:

Segmentos	Tamaño del segmento	d estimada, m	ε_t(%)
10	1.0	288.7491	0.237
20	0.5	289.2636	0.0593
50	0.2	289.4076	9.5×10^{-3}
100	0.1	289.4282	2.4×10^{-3}
200	0.05	289.4336	5.4×10^{-4}
500	0.02	289.4348	1.2×10^{-4}
1 000	0.01	289.4360	-3.0×10^{-4}
2 000	0.005	289.4369	-5.9×10^{-4}
5 000	0.002	289.4337	5.2×10^{-4}
10 000	0.001	289.4317	1.2×10^{-3}

Así, hasta cerca de 500 segmentos, la regla del trapecio de aplicación múltiple obtiene excelente precisión. Sin embargo, observe cómo el error cambia de signo y empieza a aumentar en valor absoluto más allá de los 500 segmentos. Cuando se tienen 10 000 segmentos, de hecho, parece diverger del valor verdadero. Esto se debe a la aparición del error de redondeo por el gran número de cálculos para todos esos segmentos. De esta manera, el nivel de precisión está limitado y nunca se podrá alcanzar el valor exacto de 289.4351 que se obtiene en forma analítica. Esta limitación, así como la manera de superarla se analizará con más detalle en el capítulo 22.

Del ejemplo 21.3 se llega a tres conclusiones principales:

- Para aplicaciones individuales de las funciones con buen comportamiento, la regla del trapecio de múltiples segmentos es casi exacta para el tipo de precisión requerida en diversas aplicaciones de la ingeniería.
- Si se requiere de alta exactitud, la regla del trapecio de múltiples segmentos exige un gran trabajo computacional. Aunque este trabajo resulta insignificante para una sola aplicación, puede ser muy importante cuando: *a*) se evalúan numerosas integrales, o *b*) donde la función misma es consumidora de tiempo en su evaluación. Para tales casos, quizá se requieran métodos más eficientes (serán analizados en lo que falta de este capítulo y en el próximo).
- Por último, los errores de redondeo representan una limitación en nuestra habilidad para determinar integrales. Esto se debe tanto a la precisión de la máquina como a los diversos cálculos involucrados en técnicas simples como la regla del trapecio de múltiples segmentos.

Ahora analizaremos una forma para mejorar la eficiencia. Esto es, mediante polinomios de grado superior para aproximar la integral.

21.2 REGLAS DE SIMPSON

Además de aplicar la regla del trapecio con una segmentación más fina, otra forma de obtener una estimación más exacta de una integral consiste en usar polinomios de grado superior para unir los puntos. Por ejemplo, si hay otro punto a la mitad entre $f(a)$ y $f(b)$, los tres puntos se pueden unir con una parábola (figura 21.10*a*). Si hay dos puntos igualmente espaciados entre $f(a)$ y $f(b)$, los cuatro

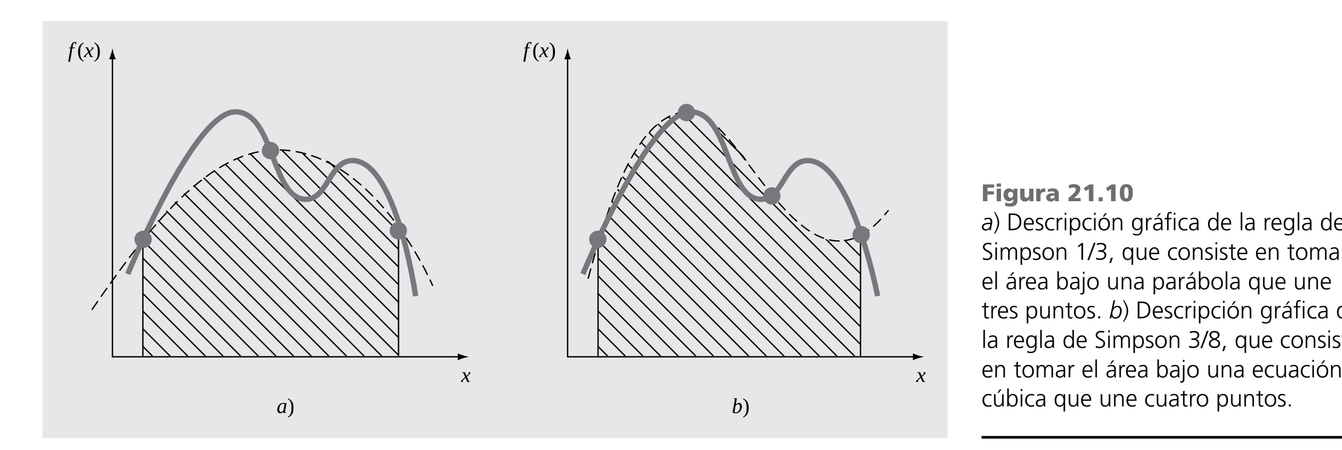


Figura 21.10
a) Descripción gráfica de la regla de Simpson 1/3, que consiste en tomar el área bajo una parábola que une tres puntos. *b*) Descripción gráfica de la regla de Simpson 3/8, que consiste en tomar el área bajo una ecuación cúbica que une cuatro puntos.

puntos se pueden unir mediante un polinomio de tercer grado (figura 21.10*b*). Las fórmulas que resultan de tomar las integrales bajo esos polinomios se conocen como *reglas de Simpson*.

21.2.1 Regla de Simpson 1/3

La regla de Simpson 1/3 resulta cuando un polinomio de interpolación de segundo grado se sustituye en la ecuación (21.1):

$$I = \int_a^b f(x)\,dx \cong \int_a^b f_2(x)\,dx$$

Si se designan a y b como x_0 y x_2, y $f_2(x)$ se representa por un polinomio de Lagrange de segundo grado [véase ecuación (18.23)], la integral se transforma en

$$I = \int_{x_0}^{x_2} \left[\frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} f(x_0) + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} f(x_1) + \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} f(x_2) \right] dx$$

Después de la integración y de las manipulaciones algebraicas, se obtiene la siguiente fórmula:

$$I \cong \frac{h}{3}[f(x_0) + 4f(x_1) + f(x_2)] \tag{21.14}$$

donde, en este caso, $h = (b - a)/2$. Esta ecuación se conoce como *regla de Simpson 1/3*, y es la segunda fórmula de integración cerrada de Newton-Cotes. La especificación “1/3” se origina del hecho de que h está dividida entre 3 en la ecuación (21.14). Una alternativa para obtenerla se muestra en el cuadro 21.3, donde se integra el polinomio de Newton-Gregory para llegar a la misma fórmula.

La regla de Simpson 1/3 también se puede expresar usando el formato de la ecuación (21.5):

$$I \cong \underbrace{(b-a)}_{\text{Ancho}} \underbrace{\frac{f(x_0) + 4f(x_1) + f(x_2)}{6}}_{\text{Altura promedio}} \tag{21.15}$$

donde $a = x_0$, $b = x_2$ y x_1 = el punto a la mitad entre a y b, que está dado por $(b + a)/2$. Observe que, de acuerdo con la ecuación (21.15), el punto medio está ponderado por dos tercios; y los dos puntos extremos, por un sexto.

Se puede demostrar que la aplicación a un solo segmento de la regla de Simpson 1/3 tiene un error de truncamiento de (cuadro 21.3)

$$E_t = -\frac{1}{90} h^5 f^{(4)}(\xi)$$

o, como $h = (b - a)/2$,

$$E_t = -\frac{(b-a)^5}{2\,880} f^{(4)}(\xi) \tag{21.16}$$

donde ξ está en algún lugar en el intervalo de a a b. Así, la regla de Simpson 1/3 es más exacta que la regla del trapecio. No obstante, una comparación con la ecuación (21.6) indica que es más exacta de lo esperado. En lugar de ser proporcional a la tercera derivada, el error es proporcional a la cuarta derivada. Esto es porque, como se muestra en el cuadro 21.3, el término del coeficiente de tercer grado se hace cero durante la integración de la interpolación polinomial. En consecuencia, la regla de Simpson 1/3 alcanza una precisión de tercer grado aun cuando se base en sólo tres puntos. En otras palabras, ¡da resultados exactos para polinomios cúbicos aun cuando se obtenga de una parábola!

Cuadro 21.3 Obtención y estimación del error de la regla de Simpson 1/3

Como se hizo en el cuadro 21.2 para la regla del trapecio, la regla de Simpson 1/3 se obtiene al integrar el polinomio de interpolación de Newton-Gregory hacia adelante (cuadro 18.2):

$$I = \int_{x_0}^{x_2} \left[f(x_0) + \Delta f(x_0)\alpha + \frac{\Delta^2 f(x_0)}{2}\alpha(\alpha-1) + \frac{\Delta^3 f(x_0)}{6}\alpha(\alpha-1)(\alpha-2) + \frac{f^{(4)}(\xi)}{24}\alpha(\alpha-1)(\alpha-2)(\alpha-3)h^4 \right] dx$$

Observe que se escribió el polinomio hasta el término de cuarto grado, en lugar de hasta el de tercer grado como se esperaría. La razón de esto se verá un poco después. Advierta también que los límites de integración van de x_0 a x_2. Por lo tanto, cuando se realizan las sustituciones para simplificar (recuerde el cuadro 21.2), la integral es de $\alpha = 0$ a 2:

$$I = h\int_{0}^{2} \left[f(x_0) + \Delta f(x_0)\alpha + \frac{\Delta^2 f(x_0)}{2}\alpha(\alpha-1) + \frac{\Delta^3 f(x_0)}{6}\alpha(\alpha-1)(\alpha-2) + \frac{f^{(4)}(\xi)}{24}\alpha(\alpha-1)(\alpha-2)(\alpha-3)h^4 \right] d\alpha$$

que al integrarse tiene

$$I = h\left[\alpha f(x_0) + \frac{\alpha^3}{2}\Delta f(x_0) + \left(\frac{\alpha^2}{6} - \frac{\alpha^2}{4}\right)\Delta^2 f(x_0) + \left(\frac{\alpha^4}{24} - \frac{\alpha^3}{6} + \frac{\alpha^2}{6}\right)\Delta^3 f(x_0) + \left(\frac{\alpha^5}{120} - \frac{\alpha^4}{16} + \frac{11\alpha^3}{72} - \frac{\alpha^2}{8}\right) f^{(4)}(\xi)h^4 \right]_0^2$$

y evaluando en los límites se obtiene

$$I = h\left[2f(x_0) + 2\Delta f(x_0) + \frac{\Delta^2 f(x_0)}{3} + (0)\Delta^3 f(x_0) - \frac{1}{90} f^{(4)}(\xi)h^4 \right] \tag{C21.3.1}$$

Observe el resultado significativo de que el coeficiente de la tercera diferencia dividida es cero. Debido a que $\Delta f(x_0) = f(x_1) - f(x_0)$ y $\Delta^2 f(x_0) = f(x_2) - 2f(x_1) + f(x_0)$, la ecuación (C21.3.1) se reescribe como

$$I = \underbrace{\frac{h}{3}[f(x_0) + 4f(x_1) + f(x_2)]}_{\text{Regla de Simpson 1/3}} - \underbrace{\frac{1}{90} f^{(4)}(\xi)h^5}_{\text{Error de truncamiento}}$$

Así, el primer término es la regla de Simpson 1/3 y el segundo es el error de truncamiento. Puesto que se suprime la tercera diferencia dividida, se obtiene el resultado significativo de que la fórmula tiene una precisión de tercer grado.

EJEMPLO 21.4 Aplicación simple de la regla de Simpson 1/3

Planteamiento del problema Con la ecuación (21.15) integre

$$f(x) = 0.2 + 25x - 200x^2 + 675x^3 - 900x^3 + 400x^5$$

desde $a = 0$ hasta $b = 0.8$. Recuerde que la integral exacta es 1.640533

Solución

$f(0) = 0.2 \qquad f(0.4) = 2.456 \qquad f(0.8) = 0.232$

Por lo tanto, la ecuación (21.15) se utiliza para calcular

$$I \cong 0.8\frac{0.2+4(2.456)+0.232}{6} = 1.367467$$

que representa un error exacto de

$$E_t = 1.640533 - 1.367467 = 0.2730667 \quad \varepsilon_t = 16.6\%$$

que es aproximadamente 5 veces más precisa que una sola aplicación de la regla del trapecio (ejemplo 21.1).

El error estimado es [ecuación (21.16)]

$$E_a = -\frac{(0.8)^5}{2\,880}(-2\,400) = 0.2730667$$

donde –2 400 es el promedio de la cuarta derivada en el intervalo, obtenida usando la ecuación (PT6.4). Como en el ejemplo 21.1, el error está aproximado (E_a), debido a que el promedio de la cuarta derivada no es una estimación exacta de $f^{(4)}(\xi)$. Sin embargo, como este caso tiene que ver con un polinomio de quinto grado, el resultado concuerda.

21.2.2 La regla de Simpson 1/3 de aplicación múltiple

Así como en la regla del trapecio, la regla de Simpson se mejora al dividir el intervalo de integración en varios segmentos de un mismo tamaño (figura 21.11):

$$h = \frac{b-a}{n} \tag{21.17}$$

La integral total se puede representar como

$$I = \int_{x_0}^{x_2} f(x)\,dx + \int_{x_2}^{x_4} f(x)\,dx + \cdots + \int_{x_{n-2}}^{x_n} f(x)\,dx$$

Al sustituir la regla de Simpson 1/3 en cada integral se obtiene

$$I \cong 2h\frac{f(x_0)+4f(x_1)+f(x_2)}{6} + 2h\frac{f(x_2)+4f(x_3)+f(x_4)}{6} + \cdots + 2h\frac{f(x_{n-2})+4f(x_{n-1})+f(x_n)}{6}$$

o, combinando términos y usando la ecuación (21.17),

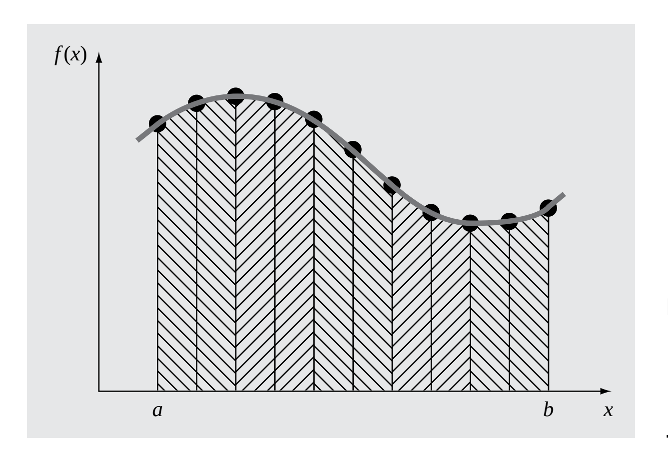


Figura 21.11
Representación gráfica de la regla de Simpson 1/3 de aplicación múltiple. Observe que el método se puede emplear sólo si el número de segmentos es par.

$$I \cong \underbrace{(b-a)}_{\text{Ancho}} \underbrace{\frac{f(x_0) + 4\sum_{i=1,3,5}^{n-1} f(x_i) + 2\sum_{j=2,4,6}^{n-2} f(x_j) + f(x_n)}{3n}}_{\text{Peso promedio}} \tag{21.18}$$

Observe que, como se ilustra en la figura 21.11, se debe utilizar un número par de segmentos para implementar el método. Además, los coeficientes "4" y "2" en la ecuación (21.18) a primera vista parecerían peculiares. No obstante, siguen en forma natural la regla de Simpson 1/3. Los puntos impares representan el término medio en cada aplicación y, por lo tanto, llevan el peso de 4 de la ecuación (21.15). Los puntos pares son comunes a aplicaciones adyacentes y, por lo tanto, se cuentan dos veces.

Un error estimado en la regla de Simpson de aplicación múltiple se obtiene de la misma forma que en la regla del trapecio: sumando los errores individuales de los segmentos y sacando el promedio de la derivada para llegar a

$$E_a = -\frac{(b-a)^5}{180n^4}\bar{f}^{(4)} \tag{21.19}$$

donde $\bar{f}^{(4)}$ es el promedio de la cuarta derivada en el intervalo.

EJEMPLO 21.5 Versión de la regla de Simpson 1/3 de aplicación múltiple

Planteamiento del problema Utilice la ecuación (21.18) con $n = 4$ para estimar la integral de

$$f(x) = 0.2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$$

desde $a = 0$ hasta $b = 0.8$. Recuerde que la integral exacta es 1.640533.

Solución $n = 4$ ($h = 0.2$):

$f(0) = 0.2$ $\quad f(0.2) = 1.288$

$f(0.4) = 2.456$ $\quad f(0.6) = 3.464$

$f(0.8) = 0.232$

A partir de la ecuación (21.18),

$$I = 0.8\frac{0.2 + 4(1.288 + 3.464) + 2(2.456) + 0.232}{12} = 1.623467$$

$$E_t = 1.640533 - 1.623467 = 0.017067 \qquad \varepsilon_t = 1.04\%$$

El error estimado [ecuación (21.19)] es

$$E_a = -\frac{(0.8)^5}{180(4)^4}(-2\,400) = 0.017067$$

El ejemplo anterior demuestra que la versión de la regla de Simpson 1/3 de aplicación múltiple da resultados muy precisos. Por esta razón, se considera mejor que la regla del trapecio en la mayoría de las aplicaciones. Sin embargo, como se indicó antes, está limitada a los casos donde los valores están equidistantes. Además, está limitada a situaciones en las que hay un número par de segmentos y un número impar de puntos. En consecuencia, como se analizará en la siguiente sección, una fórmula de segmentos impares y puntos pares, conocida como regla de Simpson 3/8, se usa junto con la regla 1/3 para permitir la evaluación de números de segmentos tanto pares como impares.

21.2.3 Regla de Simpson 3/8

De manera similar a la obtención de la regla del trapecio y Simpson 1/3, es posible ajustar un polinomio de Lagrange de tercer grado a cuatro puntos e integrarlo:

$$I = \int_a^b f(x)\,dx \cong \int_a^b f_3(x)\,dx$$

para obtener

$$I \cong \frac{3h}{8}[f(x_0) + 3f(x_1) + 3f(x_2) + f(x_3)]$$

donde $h = (b - a)/3$. Esta ecuación se llama *regla de Simpson 3/8* debido a que h se multiplica por 3/8. Ésta es la tercera fórmula de integración cerrada de Newton-Cotes. La regla 3/8 se expresa también en la forma de la ecuación (21.5):

$$I \cong \underbrace{(b-a)}_{\text{Ancho}} \underbrace{\frac{f(x_0) + 3f(x_1) + 3f(x_2) + f(x_3)}{8}}_{\text{Altura promedio}} \tag{21.20}$$

Así los dos puntos interiores tienen pesos de tres octavos, mientras que los puntos extremos tienen un peso de un octavo. La regla de Simpson 3/8 tiene un error de

$$E_t = -\frac{3}{80} h^5 f^{(4)}(\xi)$$

o, como $h = (b - a)/3$,

$$E_t = -\frac{(b-a)^5}{6\,480} f^{(4)}(\xi) \tag{21.21}$$

Puesto que el denominador de la ecuación (21.21) es mayor que el de la ecuación (21.16), la regla 3/8 es más exacta que la regla 1/3.

Por lo cumún, se prefiere la regla de Simpson 1/3, ya que alcanza una exactitud de tercer orden con tres puntos en lugar de los cuatro puntos requeridos en la versión 3/8. No obstante, la regla de 3/8 es útil cuando el número de segmentos es impar. Como ilustración, en el ejemplo 21.5 usamos la regla de Simpson para integrar la función con cuatro segmentos. Suponga que usted desea una estimación con cinco segmentos. Una opción podría ser utilizar una versión de la regla del trapecio de aplicación múltiple, como se hizo en los ejemplos 21.2 y 21.3. Quizá esto no sea recomendable, sin embargo, debido al gran error de truncamiento asociado con dicho método. Una alternativa sería aplicar la regla de Simpson 1/3 a los dos primeros segmentos y la regla de Simpson 3/8 a los últimos tres (figura 21.12). De esta forma, podríamos obtener un estimado con una exactitud de tercer orden durante todo el intervalo.

EJEMPLO 21.6 Regla de Simpson 3/8

Planteamiento del problema

a) Con la regla de Simpson 3/8 integre

$$f(x) = 0.2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$$

desde $a = 0$ hasta $b = 0.8$.

b) Úsela junto con la regla de Simpson 1/3 con la finalidad de integrar la misma función en cinco segmentos.

Solución

a) Una sola aplicación de la regla de Simpson 3/8 requiere cuatro puntos equidistantes:

$f(0) = 0.2 \qquad f(0.2667) = 1.432724$

$f(0.5333) = 3.487177 \qquad f(0.8) = 0.232$

Utilizando la ecuación (21.20),

$$I \cong 0.8\frac{0.2 + 3(1.432724 + 3.487177) + 0.232}{8} = 1.519170$$

$$E_t = 1.640533 - 1.519170 = 0.1213630 \qquad \varepsilon_t = 7.4\%$$

$$E_a = -\frac{(0.8)^5}{6\,480}(-2\,400) = 0.1213630$$

b) Los datos necesarios para una aplicación con cinco segmentos ($h = 0.16$) son

$f(0) = 0.2 \qquad f(0.16) = 1.296919$

$f(0.32) = 1.743393 \qquad f(0.48) = 3.186015$

$f(0.64) = 3.181929 \qquad f(0.80) = 0.232$

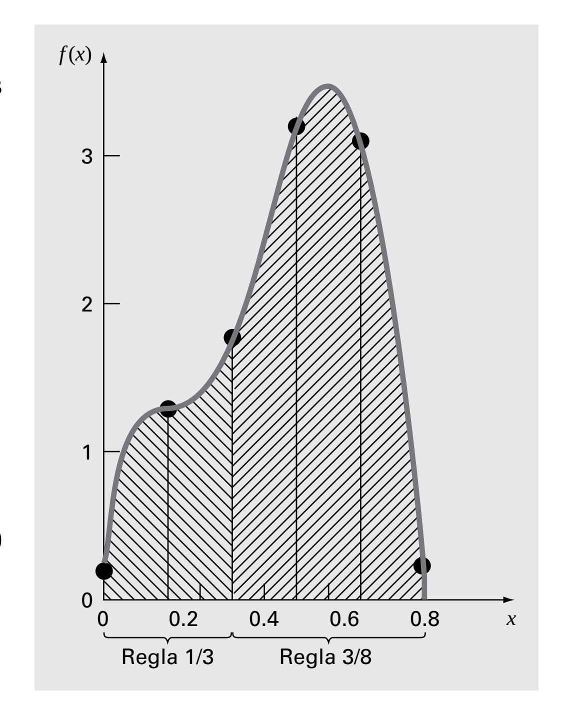


Figura 21.12
Ilustración de cómo se utilizan en conjunto las reglas de Simpson 1/3 y 3/8 para manejar aplicaciones múltiples con números impares de intervalos.

La integral para los dos primeros segmentos se obtiene usando la regla de Simpson 1/3:

$$I \cong 0.32\frac{0.2 + 4(1.296919) + 1.743393}{6} = 0.3803237$$

Para los últimos tres segmentos, la regla 3/8 se utiliza para obtener

$$I \cong 0.48\frac{1.743393 + 3(3.186015 + 3.181929) + 0.232}{8} = 1.264754$$

La integral total se calcula sumando los dos resultados:

$$I = 0.3803237 + 1.264753 = 1.645077$$

$$E_t = 1.640533 - 1.645077 = -0.00454383 \quad \varepsilon_t = -0.28\%$$

21.2.4 Algoritmos computacionales para las reglas de Simpson

En la figura 21.13 se muestran pseudocódigos para las diferentes formas de las reglas de Simpson. Observe que todas están diseñadas para datos en forma tabular. Un programa general deberá tener la capacidad de evaluar tanto las funciones como las ecuaciones conocidas. En el capítulo 22 ilustraremos cómo se manipulan las funciones.

Advierta que el programa de la figura 21.13*d* está escrito para usar un número de segmentos par o impar. En el caso par, la regla de Simpson 1/3 se aplica a cada par de segmentos y los resultados se suman para calcular la integral total. En el caso impar, la regla de Simpson 3/8 se aplica a los tres últimos segmentos; y la regla 1/3, a los segmentos anteriores.

21.2.5 Fórmulas cerradas de Newton-Cotes de grado superior

Como se observó antes, la regla del trapecio y las dos reglas de Simpson son miembros de una familia de ecuaciones de integración conocidas como fórmulas de integración cerrada de Newton-Cotes. Algunas de las fórmulas se resumen en la tabla 21.2, junto con el error de truncamiento.

a)

```
FUNCTION Simp13 (h, f0, f1, f2)
  Simp13 = 2*h*(f0+4*f1+f2) / 6
END Simp13
```

b)

```
FUNCTION Simp38 (h, f0, f1, f2, f3)
  Simp38 = 3*h* (f0+3*(f1+f2)+f3) / 8
END Simp38
```

c)

```
FUNCTION Simp13m (h,n,f)
  sum = f(0)
  DOFOR i = 1, n - 2, 2
    sum = sum + 4 * f_i + 2 * f_i+1
  END DO
  sum = sum + 4 * f_n-1 + f_n
  Simp13m = h * sum /3
END Simp13m
```

d)

```
FUNCTION SimpInt(a,b,n,f)
  h = (b - a) / n
  IF n = 1 THEN
    sum = Trap(h,f_n-1,f_n)
  ELSE
    m = n
    odd = n / 2 - INT(n / 2)
    IF odd > 0 AND n > 1 THEN
       sum = sum + Simp38(h,f_n-3,f_n-2,f_n-1,f_n)
       m = n - 3
    END IF
    IF m > 1 THEN
       sum = sum + Simp13m(h,m,f)
    END IF
  END IF
  SimpInt = sum
END SimpInt
```

Figura 21.13
Pseudocódigo para las reglas de Simpson. *a*) Regla de Simpson 1/3 para una sola aplicación, *b*) regla de Simpson 3/8 para una sola aplicación, *c*) regla de Simpson 1/3 de aplicación múltiple y *d*) regla de Simpson de aplicación múltiple para un número de segmentos tanto impares como pares. Observe que para todos los casos n debe ser ≥ 1.

Tabla 21.2 Fórmulas de integración cerrada de Newton-Cotes. Las fórmulas se presentan en el formato de la ecuación (21.5) de manera que el peso de los datos para estimar la altura promedio es aparente. El tamaño de paso está dado por $h = (b - a)/n$.

Segmentos (n)	Puntos	Nombre	Fórmula	Error de truncamiento
1	2	Regla del trapecio	$(b-a)\frac{f(x_0)+f(x_1)}{2}$	$-(1/12)h^3 f''(\xi)$
2	3	Regla de Simpson 1/3	$(b-a)\frac{f(x_0)+4f(x_1)+f(x_2)}{6}$	$-(1/90)h^5 f^{(4)}(\xi)$
3	4	Regla de Simpson 3/8	$(b-a)\frac{f(x_0)+3f(x_1)+3f(x_2)+f(x_3)}{8}$	$-(3/80)h^5 f^{(4)}(\xi)$
4	5	Regla de Boole	$(b-a)\frac{7f(x_0)+32f(x_1)+12f(x_2)+32f(x_3)+7f(x_4)}{90}$	$-(8/945)h^7 f^{(6)}(\xi)$
5	6		$(b-a)\frac{19f(x_0)+75f(x_1)+50f(x_2)+50f(x_3)+75f(x_4)+19f(x_5)}{288}$	$-(275/12\,096)h^7 f^{(6)}(\xi)$

Considere que, como en el caso de las reglas de Simpson 1/3 y 3/8, las fórmulas de cinco y seis puntos tienen el mismo orden de error. Esta característica general se satisface para fórmulas con más puntos y lleva al resultado de que las fórmulas con segmentos pares y puntos impares (por ejemplo, la regla 1/3 y la regla de Boole) usualmente son los métodos de preferencia.

Sin embargo, se debe resaltar que, en la práctica de la ingeniería, las fórmulas de grado superior (es decir, con más de cuatro puntos) son poco utilizadas. Las reglas de Simpson bastan para la mayoría de las aplicaciones. La exactitud se puede mejorar al usar la versión de aplicación múltiple. Además, cuando se conoce la función y se requiere de alta precisión, otros métodos como la integración de Romberg o la cuadratura de Gauss, descritos en el capítulo 22, ofrecen alternativas viables y atractivas.

21.3 INTEGRACIÓN CON SEGMENTOS DESIGUALES

Hasta aquí, todas las fórmulas de integración numérica se han basado en datos igualmente espaciados. En la práctica, existen muchas situaciones en donde esta suposición no se satisface y se tienen segmentos de tamaños desiguales. Por ejemplo, los datos obtenidos experimentalmente a menudo son de este tipo. En tales casos, un método consiste en aplicar la regla del trapecio a cada segmento y sumar los resultados:

$$I = h_1 \frac{f(x_0) + f(x_1)}{2} + h_2 \frac{f(x_1) + f(x_2)}{2} + \cdots + h_n \frac{f(x_{n-1}) + f(x_n)}{2} \tag{21.22}$$

donde h_i = el ancho del segmento i. Observe que éste fue el mismo procedimiento que se utilizó en la regla del trapecio de aplicación múltiple. La única diferencia entre las ecuaciones (21.8) y (21.22) es que las h en la primera son constantes. Entonces, la ecuación (21.8) podría simplificarse al agrupar términos para obtener la ecuación (21.9). Aunque esta simplificación no puede aplicarse a la ecuación (21.22), es posible desarrollar fácilmente un programa computacional para acomodar los segmentos de tamaño desigual. Antes de desarrollar este algoritmo, en el siguiente ejemplo ilustraremos cómo se aplica la ecuación (21.22) para evaluar una integral.

EJEMPLO 21.7 Regla del trapecio con segmentos desiguales

Planteamiento del problema La información de la tabla 21.3 se generó usando el mismo polinomio que se utilizó en el ejemplo 21.1. Con la ecuación (21.22) determine la integral para estos datos. Recuerde que la respuesta correcta es 1.640533.

Solución Si se aplica la ecuación (21.22) a los datos de la tabla 21.3 se obtiene

$$\begin{aligned} I &= 0.12 \frac{1.309729 + 0.2}{2} + 0.10 \frac{1.305241 + 1.309729}{2} + \cdots + 0.10 \frac{0.232 + 2.363}{2} \\ &= 0.090584 + 0.130749 + \cdots + 0.12975 = 1.594801 \end{aligned}$$

que representa un error relativo porcentual absoluto de ε_t = 2.8%.

Tabla 21.3 Datos para $f(x) = 0.2 + 25x - 200x^2 + 67x^3 - 900x^4 + 400x^5$, con valores de x desigualmente espaciados.

x	$f(x)$	x	$f(x)$
0.0	0.200000	0.44	2.842985
0.12	1.309729	0.54	3.507297
0.22	1.305241	0.64	3.181929
0.32	1.743393	0.70	2.363000
0.36	2.074903	0.80	0.232000
0.40	2.456000		

Los datos del ejemplo 21.7 se ilustran en la figura 21.14. Observe que algunos segmentos adyacentes son de la misma anchura y, en consecuencia, podrían evaluarse mediante las reglas de Simpson. Esto usualmente lleva a resultados más precisos, como lo ilustra el siguiente ejemplo.

EJEMPLO 21.8 Empleo de las reglas de Simpson en la evaluación de datos irregulares

Planteamiento del problema Vuelva a calcular la integral para los datos de la tabla 21.3, pero ahora utilice las reglas de Simpson en aquellos segmentos donde sea apropiado.

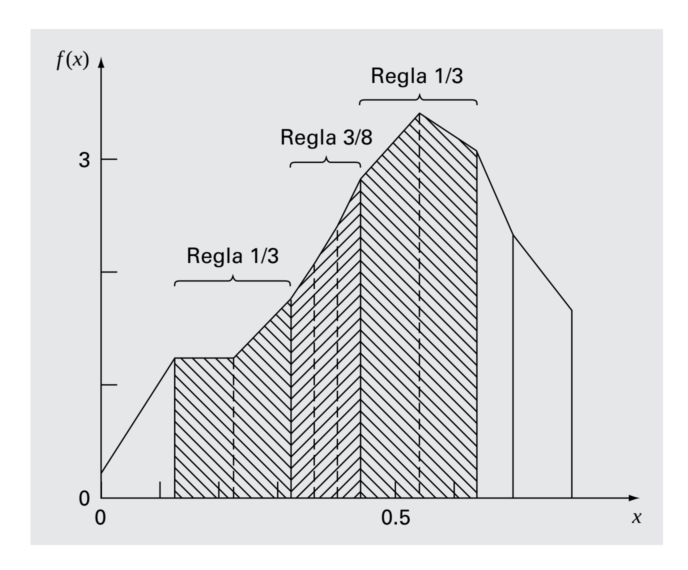


Figura 21.14
Uso de la regla del trapecio para determinar la integral de datos irregularmente espaciados. Observe cómo los segmentos sombreados podrían evaluarse con la regla de Simpson para obtener mayor precisión.

Solución El primer segmento se evalúa con la regla del trapecio:

$$I = 0.12\frac{1.309729 + 0.2}{2} = 0.09058376$$

Como los siguientes dos segmentos que van de $x = 0.12$ a 0.32 son de igual longitud, su integral se calcula con la regla de Simpson 1/3:

$$I = 0.2\frac{1.743393 + 4(1.305241) + 1.309729}{6} = 0.2758029$$

Los siguientes tres segmentos también son iguales y, por lo tanto, pueden evaluarse con la regla 3/8 para obtener $I = 0.2726863$. De manera similar, la regla 1/3 se aplica a los dos segmentos desde $x =$ 0.44 hasta 0.64 para dar $I = 0.6684701$. Finalmente, los dos últimos segmentos, que son de distinta longitud, se evalúan con la regla del trapecio para dar valores de 0.1663479 y 0.1297500, respectivamente. Se suma el área de esos segmentos individuales para tener como resultado una integral total de 1.603641. Esto representa un error de $\varepsilon_t = 2.2\%$, que es mejor al resultado que se obtuvo mediante la regla del trapecio en el ejemplo 21.7.

Programa computacional para datos irregularmente espaciados Programar la ecuación (21.22) es bastante simple. Un posible algoritmo se da en la figura 21.15*a*.

No obstante, como se demostró en el ejemplo 21.8, el procedimiento mejora si se implementan las reglas de Simpson siempre que sea posible. Por tal razón se desarrolla un segundo algoritmo que incorpora esta capacidad. Como se ilustra en la figura 21.15*b*, el algoritmo verifica la longitud de los segmentos adyacentes. Si dos segmentos consecutivos son de igual longitud, se aplica la regla de Simpson 1/3. Si tres son iguales, se utiliza la regla 3/8. Cuando los segmentos adyacentes tienen longitud desigual, se implementa la regla del trapecio.

Así, no sólo permite la evaluación de datos con segmentos desiguales, sino que al usar la información igualmente espaciada, se reduce al empleo de las reglas de Simpson. De esta manera, representa un algoritmo básico, para todo propósito en la determinación de la integral de datos tabulados.

21.4 FÓRMULAS DE INTEGRACIÓN ABIERTA

De la figura 21.3*b* recuerde que las fórmulas de integración abierta tienen límites que se extienden más allá del intervalo de los datos. La tabla 21.4 resume las fórmulas de integración abierta de Newton-Cotes. Las fórmulas se han expresado en la forma de la ecuación (21.5) de manera que los

factores de ponderación sean evidentes. Como en el caso de las versiones cerradas, pares sucesivos de las fórmulas tienen el mismo orden de error. Las fórmulas para segmentos pares y puntos impares son generalmente los métodos de preferencia, ya que requieren menos puntos para alcanzar la misma precisión que las fórmulas de segmentos impares y puntos pares.

Las fórmulas abiertas no se utilizan con frecuencia para la integración definida. No obstante, como se verá en el capítulo 22, tienen utilidad para analizar integrales impropias. Además, tendrán relevancia en nuestro análisis de los métodos de pasos múltiples, para la solución de ecuaciones diferenciales ordinarias en el capítulo 26.

a)
```
FUNCTION Trapun (x, y, n)
  LOCAL i, sum
  sum = 0
  DOFOR i = 1, n
    sum = sum + (x_i - x_i-1)*(y_i-1 + y_i)/2
  END DO
  Trapun = sum
END Trapun
```

b)
```
FUNCTION Uneven (n,x,f)
  h = x_1 - x_0
  k = 1
  sum = 0.
  DOFOR j = 1, n
    hf = x_j+1 - x_j
    IF ABS (h - hf) < .000001 THEN
       IF k = 3 THEN
         sum = sum + Simp13 (h,f_j-3,f_j-2,f_j-1)
         k = k - 1
       ELSE
         k = k + 1
        END IF
      ELSE
         IF k = 1 THEN
           sum = sum + Trap (h,f_j-1,f_j)
         ELSE
           IF k = 2 THEN
             sum = sum + Simp13 (h,f_j-2,f_j-1,f_j)
           ELSE
             sum = sum + Simp38 (h,f_j-3,f_j-2,f_j-1,f_j)
           END IF
           k = 1
         END IF
       END IF
       h = hf
     END DO
     Uneven = sum
END Uneven
```

Figura 21.15
Pseudocódigo para integrar datos desigualmente espaciados. *a*) Regla del trapecio y *b*) combinación de las reglas de Simpson y del trapecio.

Tabla 21.4 Fórmulas de integración abierta de Newton-Cotes. Las fórmulas se presentan en el formato de la ecuación (21.5), de manera que sea aparente el peso de los datos para estimar la altura promedio. El tamaño de paso está dado por $h = (b - a)/n$.

Segmentos (*n*)	Puntos	Nombre	Fórmula	Error de truncamiento
2	1	Método del punto medio	$(b-a)\, f(x_1)$	$(1/3)h^3 f''(\xi)$
3	2		$(b-a)\dfrac{f(x_1)+f(x_2)}{2}$	$(3/4)h^3 f''(\xi)$
4	3		$(b-a)\dfrac{2f(x_1)+f(x_2)+2f(x_3)}{3}$	$(14/45)h^5 f^{(4)}(\xi)$
5	4		$(b-a)\dfrac{11f(x_1)+f(x_2)+f(x_3)+11f(x_4)}{24}$	$(95/144)h^5 f^{(4)}(\xi)$
6	5		$(b-a)\dfrac{11f(x_1)+14f(x_2)+26f(x_3)+14f(x_4)+11f(x_5)}{20}$	$(41/140)h^7 f^{(6)}(\xi)$

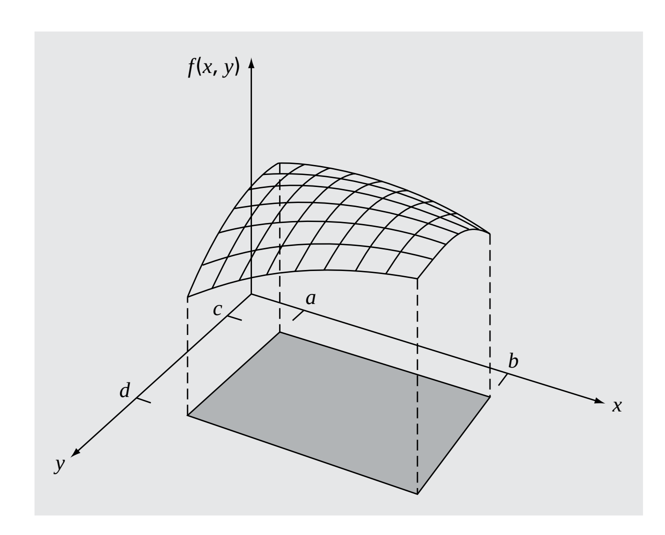


Figura 21.16
Integral doble sobre el área bajo la superficie de la función.

21.5 INTEGRALES MÚLTIPLES

Las integrales múltiples se utilizan a menudo en la ingeniería. Por ejemplo, una ecuación general para calcular el promedio de una función bidimensional puede escribirse como sigue [recuerde la ecuación (PT6.4)]:

$$\bar{f} = \frac{\int_c^d \left(\int_a^b f(x, y)\, dx \right) dy}{(d - c)(b - a)} \tag{21.23}$$

Al numerador se le llama integral doble.

Las técnicas estudiadas en este capítulo (y en el siguiente) se utilizan para evaluar integrales múltiples. Un ejemplo sencillo sería obtener la integral doble de una función sobre un área rectangular (figura 21.16).

Recuerde del cálculo que dichas integrales se pueden calcular como integrales iteradas.

$$\int_c^d \left(\int_a^b f(x, y)\, dx \right) dy = \int_a^b \left(\int_c^d f(x, y)\, dy \right) dx \tag{21.24}$$

Primero se evalúa la integral en una de las dimensiones y el resultado de esta primera integración se incorpora en la segunda dimensión. La ecuación 21.24 establece que no importa el orden de integración.

Una integral numérica doble estará basada en la misma idea. Primero se aplican métodos, como la regla de Simpson o del trapecio para segmentos múltiples, a la primera dimensión manteniendo constantes los valores de la segunda dimensión. Después, se aplica el método para integrar la segunda dimensión. El procedimiento se ilustra en el ejemplo siguiente.

EJEMPLO 21.9 Uso de la integral doble para determinar una temperatura promedio.

Planteamiento del problema Suponga que la temperatura en una placa rectangular se describe mediante la siguiente función:

$$T(x, y) = 2xy + 2x - x^2 - 2y^2 + 72$$

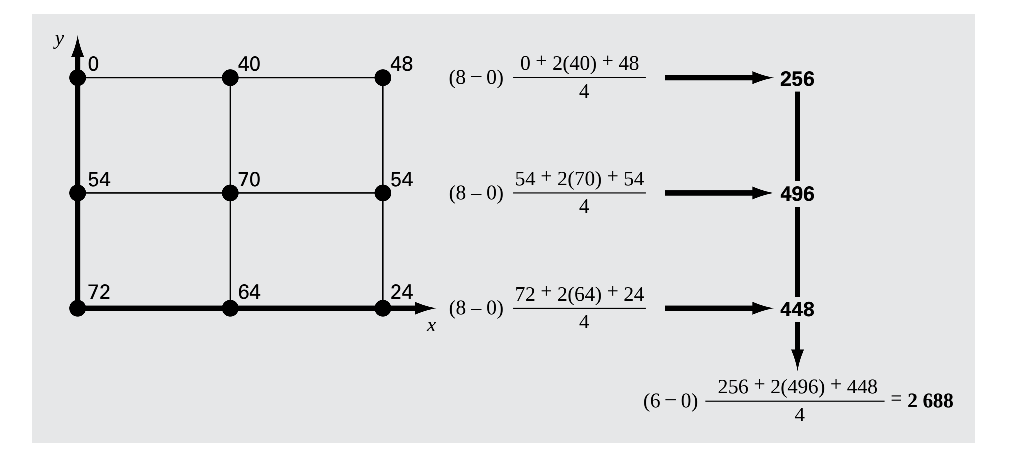


Figura 21.17
Evaluación numérica de una integral doble usando la regla del trapecio con dos segmentos.

Si la placa tiene 8 m de largo (dimensión x) y 6 m de ancho (dimensión y), calcule la temperatura promedio.

Solución Primero, se usará la regla del trapecio con dos segmentos en cada dimensión. Las temperaturas en los valores x y y necesarios se representan en la figura 21.17. Observe que un promedio simple de estos valores es 47.33. La función también se evalúa analíticamente, cuyo resultado sería 58.66667.

Para realizar numéricamente la misma evaluación se emplea primero la regla del trapecio a lo largo de la dimensión x con cada uno de los valores de y. Estos valores se integran después a lo largo de la dimensión y para dar como resultado final 2 688. Dividiendo éste entre el área se obtiene la temperatura promedio: $2\ 688/(6 \times 8) = 56$.

También podemos emplear la regla de Simpson 1/3 de la misma manera con un solo segmento. Esta integral da como resultado 2 816 y un promedio de 58.66667, que es exacto. ¿Por qué pasa esto? Recuerde que la regla de Simpson 1/3 dio resultados perfectos con polinomios cúbicos. Como el término del grado mayor en la función es de segundo grado, en el presente caso se obtiene el mismo resultado exacto.

Para funciones algebraicas de grado superior, así como con funciones trascendentes, será necesario emplear segmentos múltiples para obtener estimaciones exactas de la integral. Además, el capítulo 22 presenta técnicas más eficientes que las fórmulas de Newton-Cotes, para la evaluación de integrales de funciones dadas. Éstas con frecuencia proporcionan mejores recursos para la integración numérica de integrales múltiples.

PROBLEMAS

21.1 Evalúe la integral siguiente:

$$\int_0^{\pi/2} (6 + 3 \cos x)\, dx$$

a) en forma analítica; *b*) con una sola aplicación de la regla del trapecio; *c*) con aplicación múltiple de la regla del trapecio, con $n = 2$ y 4; *d*) con una sola aplicación de la regla de Simpson 1/3; *e*) con la aplicación múltiple de la regla de Simpson 1/3, con $n = 4$; *f*) con una sola aplicación de la regla de Simpson 3/8, y *g*) con aplicación múltiple de la regla de Simpson, con $n = 5$. Para los incisos *b*) a *g*), determine el error relativo porcentual de cada una de las estimaciones numéricas, con base en el resultado del inciso *a*).

21.2 Evalúe la integral siguiente:

$$\int_0^3 (1 - e^{-2x})\, dx$$

a) en forma analítica; *b*) con una sola aplicación de la regla del trapecio; *c*) con aplicación múltiple de la regla del trapecio, con $n = 2$ y 4; *d*) con una sola aplicación de la regla de Simpson 1/3; *e*) con aplicación múltiple de la regla de Simpson 1/3, con $n = 4$; *f*) con una sola aplicación de la regla de Simpson 3/8, y *g*) con aplicación múltiple de la regla de Simpson, con $n = 5$. Para cada una de las estimaciones numéricas de los incisos *b*) a *g*), determine el error relativo porcentual con base en el inciso *a*).

21.3 Evalúe la integral siguiente:

$$\int_{-2}^4 (1 - x - 4x^3 + 2x^5)\, dx$$

a) en forma analítica; *b*) con una sola aplicación de la regla del trapecio; *c*) con la regla del trapecio compuesta, con $n = 2$ y 4; *d*) con una sola aplicación de la regla de Simpson 1/3; *e*) con la regla de Simpson 3/8, y *f*) con la regla de Boole. Para cada una de las estimaciones numéricas de los incios *b*) a *f*), determine el error relativo porcentual con base en el inciso *a*).

21.4 Integre la función siguiente en forma analítica y con el empleo de la regla del trapecio, con $n = 1$, 2, 3 y 4:

$$\int_1^2 (x + 2/x)^2\, dx$$

Use la solución analítica para calcular los errores relativos porcentuales verdaderos para evaluar la exactitud de las aproximaciones de la regla del trapecio.

21.5 Integre la función siguiente en forma tanto analítica como con la regla de Simpson, con $n = 4$ y 5. Analice los resultados.

$$\int_{-3}^5 (4x - 3)^3\, dx$$

21.6 Integre la función siguiente tanto en forma analítica como numérica. Emplee las reglas del trapecio y de Simpson 1/3 para integrar numéricamente la función. Para ambos casos, utilice la versión de aplicación múltiple, con $n = 4$. Calcule los errores relativos porcentuales para los resultados numéricos.

$$\int_0^3 x^2 e^x\, dx$$

21.7 Integre la función siguiente tanto analítica como numéricamente. Para las evaluaciones numéricas use *a*) una sola aplicación de la regla del trapecio, *b*) la regla de Simpson 1/3, *c*) la regla de Simpson 3/8, *d*) la regla de Boole, *e*) el método del punto medio, *f*) la fórmula de integración abierta de 3 segmentos y 2 puntos, y *g*) la fórmula de integración abierta de 4 segmentos y 3 puntos. Calcule los errores relativos porcentuales de los resultados numéricos.

$$\int_0^1 14^{2x}\,dx$$

21.8 Integre la función que sigue tanto en forma analítica como numérica. Para las evaluaciones numéricas utilice *a*) una sola aplicación de la regla del trapecio; *b*) la regla de Simpson 1/3; *c*) la regla de Simpson 3/8; *d*) aplicación múltiple de reglas de Simpson, con $n = 5$; *e*) la regla de Boole; *f*) el método del punto medio; *g*) la fórmula de integración abierta de 3 segmentos y 2 puntos, y *h*) la fórmula de integración abierta de 4 segmentos y 3 puntos.

$$\int_0^3 (5+3\cos x)\,dx$$

Calcule los errores relativos porcentuales para los resultados numéricos.

21.9 Suponga que la fuerza hacia arriba de la resistencia del aire sobre un objeto que cae es proporcional al cuadrado de la velocidad. Para este caso, la velocidad se calcula con

$$v(t) = \sqrt{\frac{gm}{c_d}}\tanh\left(\sqrt{\frac{gc_d}{m}}\,t\right)$$

donde c_d = coeficiente de arrastre de segundo orden. *a*) Si $g = 9.81$ m/s², $m = 68.1$ kg y $c_d = 0.25$ kg/m, use integración analítica para determinar qué tan lejos cae el objeto en 10 segundos. *b*) Haga lo mismo, pero evalúe la integral con la regla del trapecio de segmento múltiple. Use una n suficientemente grande para obtener tres dígitos significativos de exactitud.

21.10 Evalúe la integral de los datos tabulados a continuación, con *a*) la regla del trapecio y *b*) las reglas de Simpson:

x	0	0.1	0.2	0.3	0.4	0.5
$f(x)$	1	8	4	3.5	5	1

21.11 Evalúe la integral de los datos que se tabula en seguida, con *a*) la regla del trapecio y *b*) las reglas de Simpson:

x	–2	0	2	4	6	8	10
$f(x)$	35	5	–10	2	5	3	20

21.12 Determine el valor medio de la función

$$f(x) = -46 + 45x - 14x^2 + 2x^3 - 0.075x^4$$

entre $x = 2$ y 10, por medio de *a*) graficar la función y estimar visualmente el valor medio, *b*) con la ecuación (PT6.4) y la evaluación analítica de la integral, y *c*) con la ecuación (PT6.4) y una versión de cinco segmentos de la regla de Simpson para estimar la integral. Calcule el error porcentual relativo.

21.13 La función $f(x) = 2e^{-1.5x}$ se puede utilizar para generar la tabla siguiente de datos espaciados en forma desigual:

x	0	0.05	0.15	0.25	0.35	0.475	0.6
$f(x)$	2	1.8555	1.5970	1.3746	1.1831	0.9808	0.8131

Evalúe la integral de $a = 0$ a $b = 0.6$, con el uso de *a*) medios analíticos, *b*) la regla del trapecio, y *c*) una combinación de las reglas del trapecio y de Simpson; emplee las reglas de Simpson siempre que sea posible a fin de obtener la exactitud más alta. Para los incisos *b*) y *c*), calcule el error relativo porcentual (ε_t).

21.14 Evalúe la integral doble siguiente:

$$\int_{-1}^{1}\int_0^2 (x^2 - 2y^2 + xy^3)\,dx\,dy$$

a) en forma analítica; *b*) con una aplicación múltiple de la regla del trapecio, con $n = 2$, y *c*) con aplicaciones únicas de la regla de Simpson 1/3. Para los incisos *b*) y *c*), calcule el error relativo porcentual (ε_t).

21.15 Evalúe la siguiente integral triple, *a*) en forma analítica y *b*) con el uso de aplicaciones únicas de la regla de Simpson 1/3. Para el inciso *b*) calcule el error relativo porcentual (ε_t).

$$\int_{-2}^{2}\int_0^2\int_{-3}^{1} (x^3 - 3yz)\,dx\,dy\,dz$$

21.16 Desarrolle un programa de computadora amigable para el usuario para la aplicación múltiple de la regla del trapecio, con base en la figura 21.9. Pruebe su programa con la replicación del cálculo del ejemplo 21.2.

21.17 Desarrolle un programa de cómputo amigable para el usuario para la versión de la aplicación múltiple de la regla de Simpson, con base en la figura 21.13*c*. Pruébelo con la duplicación de los cálculos del ejemplo 21.5.

21.18 Desarrolle un programa de computadora amigable para el usuario a fin de integrar datos espaciados en forma desigual, con base en la figura 21.15*b*. Pruébelo con la duplicación del cálculo del ejemplo 21.8.

21.19 Se recolectaron los siguientes datos para una sección transversal de un río (y = distancia de una ribera, H = profundidad y U = velocidad):

y, m	0	1	3	5	7	8	9	10
H, m	0	1	1.5	3	3.5	3.2	2	0
U, m/s	0	0.1	0.12	0.2	0.25	0.3	0.15	0

Use integración numérica para calcular *a*) la profundidad promedio, *b*) el área de la sección transversal, *c*) la velocidad promedio y *d*) el caudal. Observe que el área de la sección transversal (A_c) y el caudal (Q) se pueden calcular como

$$A_c = \int_0^y H(y)\,dy \qquad\qquad Q = \int_0^y H(y)U(y)\,dy$$

21.20 La concentración de salida de un reactor se mide en distintos momentos durante un periodo de 24 horas:

t, h	0	1	5.5	10	12	14	16	18	20	24
c, mg/L	1	1.5	2.3	2.1	4	5	5.5	5	3	1.2

El caudal de salida en m³/s se puede calcular con la siguiente ecuación:

$$Q(t) = 20 + 10\,\mathrm{sen}\left(\frac{2\pi}{24}(t-10)\right)$$

Use el mejor método de integración numérica para determinar el promedio ponderado de concentración de salida del reactor durante el periodo de 24 horas,

$$\bar{c} = \frac{\int_0^t Q(t)c(t)dt}{\int_0^t Q(t)dt}$$

21.21 Una viga de 11 m está sujeta a una carga, y la fuerza cortante sigue la ecuación

$$V(x) = 5 + 0.25x^2$$

donde V es la fuerza cortante y x es la distancia a lo largo de la viga. Se sabe que $V = dM/dx$ y M es el momento flexionante. La integración conduce a la relación

$$M = M_o + \int_0^x V\,dx$$

Si M_o es cero y $x = 11$, calcule M con el empleo de *a*) integración analítica, *b*) aplicación múltiple de la regla del trapecio y *c*) aplicación múltiple de las reglas de Simpson. Para los incisos *b*) y *c*) use incrementos de 1 m.

21.22 El trabajo producido por un proceso termodinámico a temperatura, presión y volumen constantes, se calcula por medio de

$$W = \int p\,dV$$

donde W es el trabajo, p la presión y V el volumen. Con el empleo de una combinación de la regla del trapecio, la de Simpson 1/3, y la de Simpson 3/8, utilice los datos siguientes para calcular el trabajo en kJ (kJ = kN · m):

Presión (kPa)	336	294.4	266.4	260.8	260.5	249.6	193.6	165.6
Volumen (m^3)	0.5	2	3	4	6	8	10	11

21.23 Determine la distancia recorrida para los datos siguientes:

t, min	1	2	3.25	4.5	6	7	8	9	9.5	10
v, m/s	5	6	5.5	7	8.5	8	6	7	7	5

a) Use la regla del trapecio, *b*) la mejor combinación de las reglas del trapecio y de Simpson, y *c*) la integración analítica de polinomios de segundo y tercer grados, determinados por regresión.

21.24 La masa total de una barra de densidad variable está dada por

$$m = \int_0^L \rho(x) A_c(x)\,dx$$

donde m = masa, $\rho(x)$ = densidad, $A_c(x)$ = área de la sección transversal, x = distancia a lo largo de la barra y L = longitud total de la barra. Se midieron los datos siguientes para una barra de 10 m de longitud. Determine la masa en kilogramos con la exactitud mejor posible.

x, m	0	2	3	4	6	8	10
ρ, g/cm^3	4.00	3.95	3.89	3.80	3.60	3.41	3.30
A_c, cm^2	100	103	106	110	120	133	150

21.25 Un estudio de ingeniería del transporte requiere que usted determine el número de autos que pasan por una intersección cuando viajan durante la hora pico de la mañana. Usted se para al lado de la carretera y cuenta el número de autos que pasan cada cuatro minutos a varias horas, como se muestra en la tabla a continuación. Utilice el mejor método numérico para determinar *a*) el número total de autos que pasan entre las 7:30 y las 9:15, y *b*) la tasa de autos que cruzan la intersección por minuto. (*Sugerencia:* Tenga cuidado con las unidades.)

Tiempo (h)	7:30	7:45	8:00	8:15	8:45	9:15
Tasa (autos por 4 min)	18	24	26	20	18	9

21.26 Determine el valor promedio para los datos de la figura P21.26. Realice la integral que se necesita para el promedio en el orden que muestra la ecuación siguiente:

$$I = \int_{x_0}^{x_n} \left[\int_{y_0}^{y_m} f(x, y)\,dy \right] dx$$

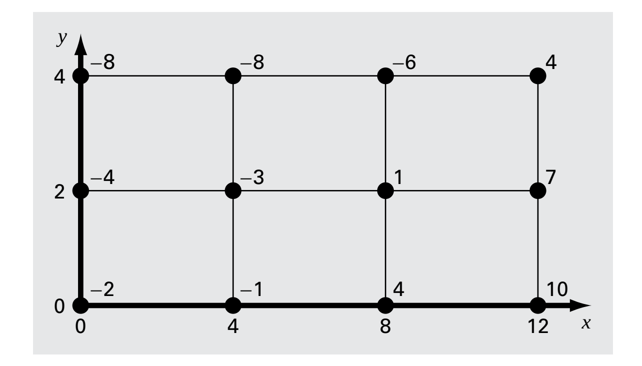


Figura P21.26

CAPÍTULO 22

Integración de ecuaciones

En la introducción de la parte seis destacamos que las funciones que vayan a integrarse de manera numérica son principalmente de dos tipos: una tabla de valores o una función. La forma de los datos tiene una influencia importante en los procedimientos que se utilizan para evaluar la integral. Con información tabulada, se está limitando al número de puntos que se tengan. En cambio, si se tiene la función, se pueden generar tantos valores de $f(x)$ como se requieran para alcanzar una exactitud aceptable (recuerde la figura PT6.7).

Este capítulo se ocupa de tres técnicas expresamente diseñadas para analizar los casos donde se tiene la función. Cada una aprovecha la posibilidad de generar valores de la función para desarrollar esquemas eficientes para la integración numérica. La primera se basa en la *extrapolación de Richardson*, que es un método que combina dos estimaciones numéricas de la integral para obtener un tercer valor más exacto. El algoritmo computacional para implementar de manera muy eficiente la extrapolación de Richardson se llama *integración de Romberg.* Esta técnica es recursiva y se utiliza para generar una estimación de la integral dentro de una tolerancia de error preespecificada.

El segundo método, la *integración adaptiva*, se basa en la división del intervalo de integración en subintervalos sucesivos más refinados de manera recursiva. Así, se utiliza espaciamiento más refinado donde la función varía más rápidamente, y espaciamiento más grueso donde la función varía lentamente, con objeto de obtener la exactitud global deseada con el menor esfuerzo de cálculo.

El tercer método se denomina *cuadratura de Gauss*. Recuerde que en el último capítulo los valores de $f(x)$ en las fórmulas de Newton-Cotes se determinaron para valores específicos de x. Por ejemplo, si se utiliza la regla del trapecio para determinar una integral, estamos restringidos a tomar el promedio ponderado de $f(x)$ en los extremos del intervalo. Las fórmulas de cuadratura de Gauss emplean valores de x que están entre a y b, de forma que resulta una estimación mucho más exacta de la integral.

Además de estas dos técnicas estándar, dedicamos una sección final a la evaluación de *integrales impropias*. En este análisis nos concentraremos en integrales con límites infinitos y en mostrar cómo un cambio de variable y de fórmulas de integración abierta son útiles en tales casos.

22.1 ALGORITMOS DE NEWTON-COTES PARA ECUACIONES

En el capítulo 21 presentamos algoritmos para versiones de aplicación múltiple de la regla del trapecio y de las reglas de Simpson. Aunque estos pseudocódigos pueden usarse para analizar ecuaciones, en nuestro esfuerzo por hacerlos compatibles tanto con los datos como con las funciones, no pueden aprovechar la ventaja de estas últimas.

La figura 22.1 muestra los pseudocódigos que están expresamente diseñados para casos donde la función es analítica. En particular, observe que ni los valores de la variable independiente, ni de la dependiente se pasan a la función por medio de su argumento, como fue el caso para los códigos del capítulo 21. Para la variable independiente x, se da el intervalo de integración (a, b) y el número de segmentos. Esta información se emplea después para generar valores igualmente espaciados de x dentro de la función. Para la variable dependiente, los valores de la función en la figura 22.1 se calculan llamando a la función que está analizándose, $f(x)$.

Desarrollamos programas de precisión simple, basados en esos pseudocódigos, para analizar el trabajo implicado y los errores en que se incurre al usar progresivamente más segmentos para estimar la integral de una función simple. Para una función analítica, las ecuaciones del error [ecuaciones (21.13) y (21.19)] indican que el aumento en el número de segmentos n resultará en estimaciones más exactas de la integral. Esta observación es justificada en la figura 22.2, la cual es una gráfica del error

a)

```
FUNCTION TrapEq (n, a, b)
  h = (b − a) / n
  x = a
  sum = f(x)
  DOFOR i = 1, n − 1
    x = x + h
    sum = sum + 2 * f(x)
  END DO
  sum = sum + f(b)
  TrapEq = (b − a) * sum / (2 * n)
END TrapEq
```

b)

```
FUNCTION SimpEq (n, a, b)
  h = (b − a) / n
  x = a
  sum = f(x)
  DOFOR i = 1, n − 2, 2
    x = x + h
    sum = sum + 4 * f(x)
    x = x + h
    sum = sum + 2 * f(x)
  END DO
  x = x + h
  sum = sum + 4 * f(x)
  sum = sum + f(b)
  SimpEq = (b − a) * sum/(3 * n)
END SimpEq
```

Figura 22.1
Algoritmos de las reglas
a) del trapecio y *b*) de Simpson 1/3 de aplicaciones múltiples donde se tiene la función.

verdadero contra n para la integral de $f(x) = 0.2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$. Observe cómo el error disminuye conforme n se incrementa. Sin embargo, note también que para grandes valores de n, el error empieza a aumentar conforme los errores de redondeo empiezan a dominar. Observe además que se requiere un número muy grande de evaluaciones de la función (y, por lo tanto, de más trabajo de cálculo) para alcanzar altos niveles de precisión. Como una consecuencia de estas desventajas, la regla del trapecio y las reglas de Simpson de aplicación múltiple algunas veces resultan inadecuadas para resolver problemas en contextos donde se necesitan alta eficiencia y errores mínimos.

22.2 INTEGRACIÓN DE ROMBERG

La *integración de Romberg* es una técnica diseñada para obtener integrales numéricas de funciones de manera eficiente. Es muy parecida a las técnicas analizadas en el capítulo 21, en el sentido de que se basa en aplicaciones sucesivas de la regla del trapecio. Sin embargo, a través de las manipulaciones matemáticas, se alcanzan mejores resultados con menos trabajo.

22.2.1 Extrapolación de Richardson

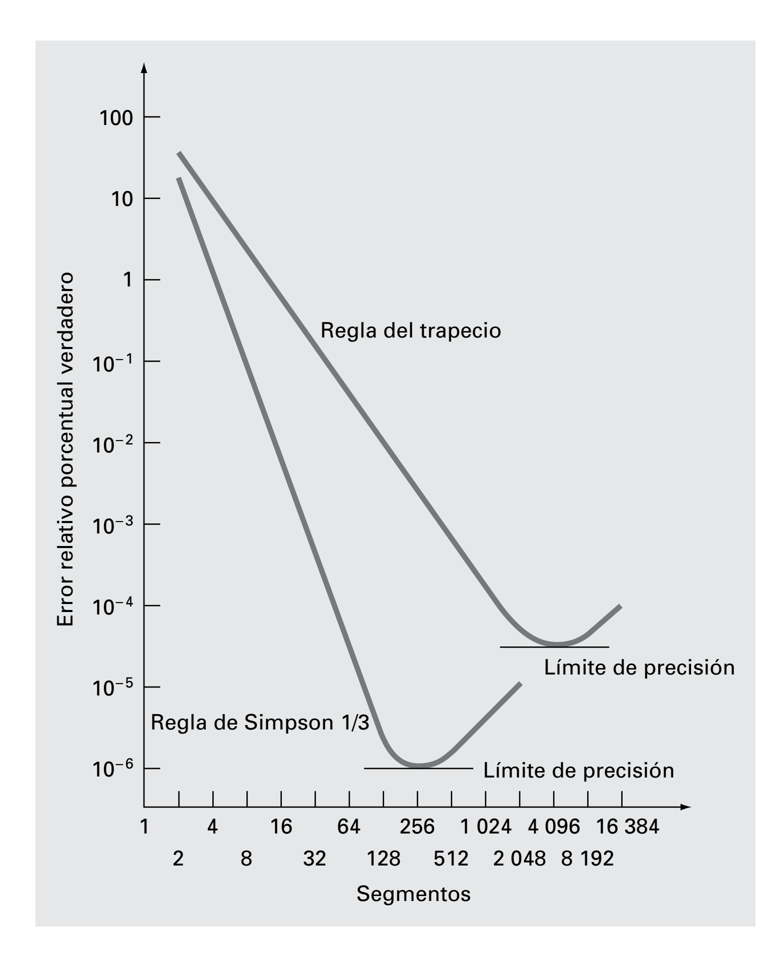
Recuerde que en la sección 10.3.3 usamos refinamiento iterativo para mejorar la solución de un conjunto de ecuaciones lineales simultáneas. Hay técnicas de corrección del error para mejorar los resultados de la integración numérica con base en la misma estimación de la integral. Dichos métodos usan dos estimaciones de una integral para calcular una tercera más exacta y, en general, se les conoce como *extrapolación de Richardson*.

La estimación y el error correspondiente a la regla del trapecio de aplicación múltiple se representa de manera general como:

$$I = I(h) + E(h)$$

donde I = el valor exacto de la integral, $I(h)$ = la aproximación obtenida de una aplicación con n segmentos de la regla del trapecio, con un tamaño de paso $h = (b - a)/n$, y $E(h)$ = el error de truncamiento. Si hacemos, por separado, dos estimaciones usando tamaños de paso h_1 y h_2 y tenemos valores exactos del error,

$$I(h_1) + E(h_1) = I(h_2) + E(h_2) \tag{22.1}$$



Ahora recuerde que el error de la regla del trapecio de aplicación múltiple puede representarse en forma aproximada mediante la ecuación (21.13) [con $n = (b - a)/h$]:

$$E \cong -\frac{b-a}{12} h^2 \bar{f}'' \tag{22.2}$$

Si se supone que $\bar{f}''$ es constante para todo tamaño de paso, la ecuación 22.2 se utiliza para determinar la razón entre los dos errores, que será

$$\frac{E(h_1)}{E(h_2)} \cong \frac{h_1^2}{h_2^2} \tag{22.3}$$

Figura 22.2
Valor absoluto del error relativo porcentual verdadero contra el número de segmentos para la determinación de la integral de $f(x) = 0.2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$, evaluada desde $a = 0$ hasta $b = 0.8$ mediante la regla del trapecio de aplicación múltiple y la regla de Simpson 1/3 de aplicación múltiple. Observe que ambos resultados indican que para un gran número de segmentos, los errores de redondeo limitan la precisión.

Este cálculo tiene el importante efecto de eliminar el término $\bar{f}''$ de los cálculos. Al hacerlo, fue posible utilizar la información contenida en la ecuación (22.2) sin un conocimiento previo de la segunda derivada de la función. Para lograr esto, se reordena la ecuación (22.3) para dar

$$E(h_1) \cong E(h_2)\left(\frac{h_1}{h_2}\right)^2$$

que se puede sustituir en la ecuación (22.1):

$$I(h_1) + E(h_2)\left(\frac{h_1}{h_2}\right)^2 \cong I(h_2) + E(h_2)$$

de donde se despeja

$$E(h_2) \cong \frac{I(h_1) - I(h_2)}{1 - (h_1/h_2)^2}$$

Así, hemos desarrollado un estimado del error de truncamiento en términos de las estimaciones de la integral y de sus tamaños de paso. La estimación se sustituye después en

$$I = I(h_2) + E(h_2)$$

para obtener una mejor estimación de la integral:

$$I \cong I(h_2) + \frac{1}{(h_1/h_2)^2 - 1}[I(h_2) - I(h_1)] \tag{22.4}$$

Se puede demostrar (Ralston y Rabinowitz, 1978) que el error de esta estimación es $O(h^4)$. Así, hemos combinado dos estimaciones con la regla del trapecio de $O(h^2)$ para obtener una nueva estimación de $O(h^4)$. En el caso especial donde el intervalo es dividido a la mitad ($h_2 = h_1/2$), esta ecuación se convierte en

$$I \cong I(h_2) + \frac{1}{2^2 - 1}[I(h_2) - I(h_1)]$$

o, agrupando términos,

$$I \cong \frac{4}{3} I(h_2) - \frac{1}{3} I(h_1) \qquad (22.5)$$

EJEMPLO 22.1 Correcciones del error en la regla del trapecio

Planteamiento del problema En el capítulo anterior (ejemplo 21.1 y tabla 21.1) empleamos varios métodos de integración numérica para evaluar la integral de $f(x) = 0.2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$ desde $a = 0$ hasta $b = 0.8$. Por ejemplo, las aplicaciones simples y múltiples de la regla del trapecio dieron los siguientes resultados:

Segmentos	h	Integral	ε_t %
1	0.8	0.1728	89.5
2	0.4	1.0688	34.9
4	0.2	1.4848	9.5

Use esta información junto con la ecuación (22.5) para calcular mejores estimaciones de la integral.

Solución Si se combinan las estimaciones con uno y dos segmentos resulta:

$$I \cong \frac{4}{3}(1.0688) - \frac{1}{3}(0.1728) = 1.367467$$

El error de la integral mejorada es $E_t = 1.640533 - 1.367467 = 0.273067$ ($\varepsilon_t = 16.6\%$), que es mejor al de las estimaciones sobre las que se basó.

De la misma manera, las estimaciones con dos y cuatro segmentos se combinan para obtener

$$I \cong \frac{4}{3}(1.4848) - \frac{1}{3}(1.0688) = 1.623467$$

que representa un error $E_t = 1.640533 - 1.623467 = 0.017067$ ($\varepsilon_t = 1.0\%$).

La ecuación (22.4) proporciona una forma de combinar dos aplicaciones de la regla del trapecio con un error $O(h^2)$, para calcular una tercera estimación con un error $O(h^4)$. Este procedimiento es un subconjunto de un método más general para combinar integrales y obtener mejores estimaciones. Así, en el ejemplo 22.1, calculamos dos integrales mejoradas de $O(h^4)$ con base en tres estimaciones con la regla del trapecio. Estos dos cálculos mejorados pueden, a su vez, combinarse para generar un valor aún mejor con $O(h^6)$. En el caso especial donde las estimaciones originales con la regla del trapecio se basan en la división sucesiva de la mitad del tamaño de paso, la ecuación usada para una exactitud $O(h^6)$ es

$$I \cong \frac{16}{15} I_m - \frac{1}{15} I_l \qquad (22.6)$$

donde I_m e I_l son las estimaciones mayor y menor, respectivamente. De manera similar, dos resultados $O(h^6)$ se combinan para calcular una integral que es $O(h^8)$ utilizando

$$I \cong \frac{64}{63} I_m - \frac{1}{63} I_l \tag{22.7}$$

EJEMPLO 22.2 Corrección del error de orden superior en estimaciones de la integral

Planteamiento del problema En el ejemplo 22.1 usamos la extrapolación de Richardson para calcular dos estimaciones de la integral de $O(h^4)$. Utilice la ecuación (22.6) para combinar esas estimaciones y calcular una integral con $O(h^6)$.

Solución Las dos estimaciones de la integral de $O(h^4)$ obtenidas en el ejemplo 22.1 fueron 1.367467 y 1.623467. Se sustituyen tales valores en la ecuación (22.6) y se obtiene

$$I = \frac{16}{15}(1.623467) - \frac{1}{15}(1.367467) = 1.640533$$

que es el resultado correcto hasta siete cifras significativas que se utilizaron en este ejemplo.

22.2.2 El algoritmo de integración de Romberg

Observe que los coeficientes en cada una de las ecuaciones de extrapolación [ecuaciones (22.5), (22.6) y (22.7)] suman 1. De esta manera, representan factores de ponderación que, conforme aumenta la exactitud, dan un peso relativamente mayor a la mejor estimación de la integral. Estas formulaciones se expresan en una forma general muy adecuada para la implementación en computadora:

$$I_{j,k} \cong \frac{4^{k-1} I_{j+1,k-1} - I_{j,k-1}}{4^{k-1} - 1} \tag{22.8}$$

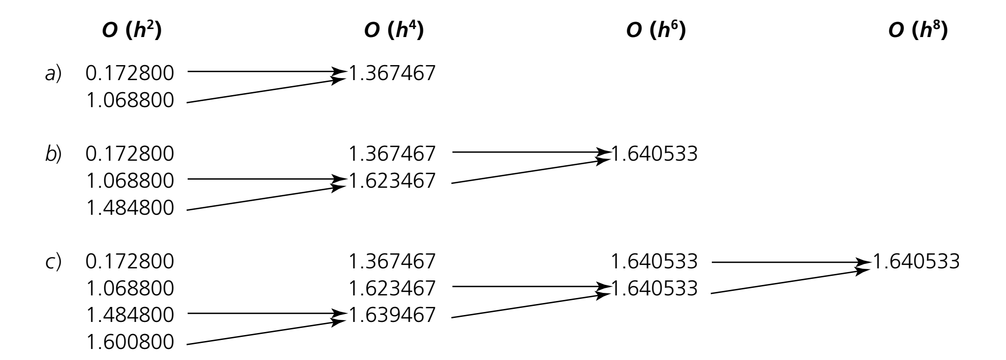


Figura 22.3
Representación gráfica de la sucesión de estimaciones de la integral generadas usando la integración de Romberg. *a*) Primera iteración. *b*) Segunda iteración. *c*) Tercera iteración.

donde $I_{j+1,k-1}$ e $I_{j,k-1}$ = las integrales más y menos exactas, respectivamente; e $I_{j,k}$ = la integral mejorada. El subíndice k significa el nivel de la integración, donde $k = 1$ corresponde a la estimación original con la regla del trapecio, $k = 2$ corresponde a $O(h^4)$, $k = 3$ a $O(h^6)$, y así sucesivamente. El subíndice j se usa para distinguir entre las estimaciones más ($j + 1$) y menos (j) exactas. Por ejemplo, con $k = 2$ y $j = 1$, la ecuación (22.8) se convierte en

$$I_{1,2} \cong \frac{4I_{2,1} - I_{1,1}}{3}$$

que es equivalente a la ecuación (22.5).

La forma general representada por la ecuación (22.8) se atribuye a Romberg, y su aplicación sistemática para evaluar integrales se denomina *integración de Romberg.* La figura 22.3 es una representación gráfica de la sucesión de estimaciones de la integral generadas usando este procedimiento. Cada matriz corresponde a una sola iteración. La primera columna contiene las evaluaciones de la regla del trapecio, designadas por $I_{j,1}$, donde $j = 1$ indica una aplicación con un solo segmento (el tamaño de paso es $b - a$), $j = 2$ corresponde a una aplicación con dos segmentos [el tamaño de paso es $(b - a)/2$], $j = 3$ corresponde a una aplicación de cuatro segmentos [el tamaño de paso es $(b - a)/4$], y así sucesivamente. Las otras columnas de la matriz se generan mediante la aplicación sistemática de la ecuación (22.8) para obtener sucesivamente mejores estimaciones de la integral.

Por ejemplo, la primera iteración (figura 22.3*a*) consiste en calcular las estimaciones con la regla del trapecio para uno y dos segmentos ($I_{1,1}$ e $I_{2,1}$). La ecuación (22.8) se utiliza después para calcular el elemento $I_{1,2} = 1.367467$, el cual tiene un error de $O(h^4)$.

Ahora, debemos verificar y establecer si este resultado es adecuado a nuestras necesidades. Como en los otros métodos aproximados de este libro, se requiere un criterio de paro, o de terminación, para evaluar la exactitud de los resultados. Un método que es útil para el propósito actual es [ecuación (3.5)]

$$|\varepsilon_a| = \left|\frac{I_{1,k} - I_{1,k-1}}{I_{1,k}}\right| 100\% \tag{22.9}$$

donde ε_a = una estimación del error relativo porcentual. De esta manera, como sucedió en los anteriores procesos iterativos, se compara la nueva estimación con un valor anterior. Cuando la diferencia entre los valores anteriores y nuevos representada por ε_a está por debajo de un criterio de error preespecificado ε_s, termina el cálculo. En la figura 22.3*a*, esta evaluación indica 21.8% de cambio con respecto a la primera iteración.

El objetivo de la segunda iteración (figura 22.3*b*) es obtener la estimación $O(h^6)$, $I_{1,3}$. Para hacerlo, se determina una estimación más con la regla del trapecio, $I_{3,1} = 1.4848$. Después, ésta se combina con $I_{2,1}$ usando la ecuación (22.8) para generar $I_{2,2} = 1.623467$. El resultado se combina, a su vez, con $I_{1,2}$ para obtener $I_{1,3} = 1.640533$. Se aplica la ecuación (22.9) para determinar que este resultado representa un cambio del 1.0% cuando se compara con el resultado previo, $I_{1,2}$.

La tercera iteración (figura 22.3*c*) continúa con el proceso de la misma forma. En tal caso, la estimación del trapecio se suma a la primera columna y, después, se aplica la ecuación (22.8) para calcular en forma sucesiva integrales más exactas a lo largo de la diagonal inferior. Después de sólo tres iteraciones, debido a que se evalúa un polinomio de quinto grado, el resultado ($I_{1,4} = 1.640533$) es exacto.

La integración de Romberg es más eficiente que las reglas del trapecio y de Simpson analizadas en el capítulo 21. Por ejemplo, para la determinación de una integral como la de la figura 22.1, la regla de Simpson 1/3 requeriría una aplicación con 256 segmentos para dar un estimado de 1.640533. Aproximaciones más finas no serán posibles debido al error de redondeo. En cambio, la integración de Romberg ofrece un resultado exacto (a siete cifras significativas) basada en la combinación de aplicaciones de la regla del trapecio con uno, dos, cuatro y ocho segmentos; es decir, ¡con sólo 15 evaluaciones de la función!

La figura 22.4 representa el pseudocódigo para la integración de Romberg. Mediante el uso de ciclos, este algoritmo implementa el método en una forma eficiente. La integración de Romberg está diseñada para casos en donde se conoce la función que se va a integrar. Esto se debe a que el conocimiento de la función permite las evaluaciones requeridas para la implementación inicial de la regla del trapecio. Los datos tabulados rara vez están en la forma requerida para dividirlos a la mitad sucesivamente como se requiere.

```
FUNCTION Romberg (a, b, maxit, es)
  LOCAL I(10, 10)
  n = 1
  I1,1 = TrapEq(n, a, b)
  iter = 0
  DO
    iter = iter + 1
    n = 2^iter
    Iiter+1,1 = TrapEq(n, a, b)
    DOFOR k = 2, iter + 1
      j = 2 + iter - k
      Ij,k = (4^(k-1) * Ij+1,k-1 - Ij,k-1)/(4^(k-1) - 1)
    END DO
    ea = ABS((I1,iter+1 - I2,iter)/I1,iter+1) * 100
    IF (iter ≥ maxit OR ea ≤ es) EXIT
  END DO
  Romberg = I1,iter+1
END Romberg
```

Figura 22.4
Pseudocódigo para la integración de Romberg, que usa la versión de segmentos del mismo tamaño de la regla del trapecio, a partir de la figura 22.1.

22.3 CUADRATURA ADAPTIVA

Aunque la integración de Romberg es más eficiente que la regla compuesta de Simpson 1/3, ambas usan puntos igualmente espaciados. La perspectiva global ignora el hecho de que muchas funciones tienen regiones de alta variabilidad junto con otras secciones en las que el cambio es gradual.

Los métodos de *cuadratura adaptiva* resuelven esta situación ajustando el tamaño de paso de modo que se usan pequeños intervalos en regiones de variación rápida, e intervalos mayores cuya función cambia gradualmente. La mayoría de estas técnicas se basa en la aplicación de la regla de Simpson 1/3 a subintervalos de un modo muy similar a la utilización de la regla de compuesto trapezoidal en la extrapolación de Richardson. Es decir, la regla de 1/3 se aplica en dos niveles de refinamiento, y la diferencia entre estos dos niveles se usa para estimar el error de truncamiento. Si el error de truncamiento es aceptable, no se necesita más refinamiento, y el estimado de la integral para el subintervalo se considera aceptable. Si el estimado de error es demasiado grande, entonces se refina el tamaño de paso y se repite el proceso hasta que el error caiga a niveles aceptables. Luego se calcula la integral total como la suma de los estimados de integrales para los subintervalos.

La base teórica de este enfoque se puede ilustrar para un intervalo $x = a$ a $x = b$ con una amplitud de $h_1 = b - a$. Se puede hacer un primer estimado de la integral con la regla de Simpson 1/3,

$$I(h_1) = \frac{h_1}{6}(f(a) + 4f(c) + f(b)) \tag{22.10}$$

donde $c = (a + b)/2$.

Como en la extrapolación de Richardson, se puede obtener un estimado más refinado disminuyendo a la mitad el tamaño de paso. Es decir, aplicando la regla de Simpson 1/3 de aplicación múltiple con $n = 4$.

$$I(h_2) = \frac{h_2}{6}(f(a) + 4f(d) + 2f(c) + 4f(e) + f(b)) \tag{22.11}$$

donde $d = (a + c)/2$, $e = (c + b)/2$ y $h_2 = h_1/2$.

Como $I(h_1)$ e $I(h_2)$ son estimados de la misma integral, su diferencia da una medida del error. Es decir,

$$E \cong I(h_2) - I(h_1) \tag{22.12}$$

Además, el estimado y el error correspondiente a cualquiera de las dos aplicaciones se puede representar generalmente como

$$I = I(h) + E(h) \tag{22.13}$$

donde I = valor exacto de la integral, $I(h)$ = aproximación de una aplicación de n segmentos de la regla de Simpson 1/3 con tamaño de paso $h = (b - a)/n$ y $E(h)$ = el error correspondiente de truncamiento.

Usando un enfoque similar a la extrapolación de Richardson, se puede obtener un estimado del error del estimado más refinado, $I(h_2)$, como función de la diferencia entre los dos estimados de integral.

$$E(h_2) = \frac{1}{15}[I(h_2) - I(h_1)] \tag{22.14}$$

El error se puede entonces agregar a $I(h_2)$ para generar un estimado todavía mejor

$$I = I(h_2) + \frac{1}{15}[I(h_2) - I(h_1)] \tag{22.15}$$

Este resultado es equivalente a la *regla de Boole*.

Las ecuaciones anteriores desarrolladas se pueden ahora combinar en un algoritmo eficiente. La figura 22.5 presenta un pseudocódigo para un algoritmo, así que está basado en un archivo M de MATLAB desarrollado por Cleve Moler (2005).

```
FUNCTION quadapt(a, b)              (función principal de llamada)
tol = 0.000001
c = (a + b)/2                                    (inicialización)
fa = f(a)
fc = f(c)
fb = f(b)
quadapt = qstep(a, b, tol, fa, fc, fb)
END quadapt

FUNCTION qstep(a, b, tol, fa, fc, fb) (función recursiva)
h1 = b − a
h2 = h1/2
c = (a + b)/2
fd = f((a + c)/2)
fe = f((c + b)/2)
I1 = h1/6 * (fa + 4 * fc + fb)            (Regla de Simpson 1/3)
I2 = h2/6 * (fa + 4 * fd + 2 * fc + 4 * fe + fb)
IF |I2 − I1| ≤ tol THEN     (terminar después de la regla de Boole)
  I = I2 + (I2 − I1)/15
ELSE                          (llamadas recursivas si se necesitan)
  Ia = qstep(a, c, tol, fa, fd, fc)
  Ib = qstep(c, b, tol, fc, fe, fb)
  I = Ia + Ib
END IF
qstep = I
END qstep
```

Figura 22.5
Pseudocódigo para algoritmo de cuadratura adaptiva, basado en un archivo M de MATLAB desarrollado por Moler (2005).

La función consiste en una función principal de llamada, `quadapt`, junto con una función recursiva, `qstep`, que realmente ejecuta la integración. Como se muestra en la figura 22.5, tanto `quadapt` como `qstep` deben tener acceso a otra función, `f`, que evalúa el integrando.

La función principal de llamada, `quadapt`, acepta como argumentos de entrada a los límites de integración *a* y *b*. Después de establecer la tolerancia, se calculan las evaluaciones de la función que se necesita para la aplicación inicial de la regla de Simpson 1/3 [ecuación (22.10)]. Estos valores, junto con los límites de integración, se pasan luego a `qstep`. Dentro de `qstep` se determinan los restantes tamaños de paso y valores de función, y se calculan los dos estimados de integral [ecuaciones (22.10) y (22.11)].

En este punto, se estima el error como la diferencia absoluta entre los estimados de integral. Dependiendo del valor del error, pueden suceder dos cosas:

1. Si el error es menor o igual que la tolerancia, se genera la regla de Boole, la función termina y se entrega el resultado.
2. Si el error es mayor que la tolerancia, se llama dos veces `qstep` para evaluar cada uno de los dos subintervalos de la llamada actual.

Las dos llamadas recursivas en el segundo paso representan la belleza real de este algoritmo. Estas llamadas siguen subdividiendo hasta que se cumple con la tolerancia. Una vez que esto sucede, sus resultados se pasan de regreso por la ruta recursiva, combinándose en el camino con los otros estimados de integral. El proceso termina cuando se satisface la llamada final, se evalúa la integral total y se devuelve a la función principal de llamada.

Se debe destacar que el algoritmo de la figura 22.5 es una versión simplificada de la función `quad`, que es la función profesional para ubicar raíces utilizada en MATLAB. Por lo tanto, no protege contra fallas como casos en los que las integrales no existen. Sin embargo, funciona bien para muchas aplicaciones, y ciertamente para ilustrar cómo opera la cuadratura adaptiva.

22.4 CUADRATURA DE GAUSS

En el capítulo 21 estudiamos fórmulas de integración numérica o cuadratura conocidas como ecuaciones de Newton-Cotes. Una característica de estas fórmulas (con excepción del caso especial de la sección 21.3) fue que la estimación de la integral se basó en valores igualmente espaciados de la función. En consecuencia, la localización de los puntos que se usaron en estas ecuaciones eran predeterminados o fijos.

Por ejemplo, como se describe en la figura 22.6*a*, la regla del trapecio se basa en obtener el área bajo la línea recta que une los valores de la función, en los extremos del intervalo de integración. La fórmula que se utiliza para calcular esta área es

$$I \cong (b-a)\frac{f(a)+f(b)}{2} \tag{22.16}$$

donde a y b son los límites de integración y $b - a$ = el ancho del intervalo de integración. Debido a que la regla del trapecio necesita los puntos extremos, existen casos como el de la figura 22.6*a*, donde la fórmula puede dar un gran error.

Ahora, suponga que se elimina la restricción de los puntos fijos y se tuviera la libertad de evaluar el área bajo una línea recta que uniera dos puntos cualesquiera de la curva. Al ubicar esos puntos en forma inteligente, definiríamos una línea recta que equilibrara los errores negativo y positivo. Así que, como en la figura 22.6*b*, llegaríamos a una mejor estimación de la integral.

Cuadratura de Gauss es el nombre de una clase de técnicas para realizar tal estrategia. Las fórmulas particulares de cuadratura de Gauss descritas en esta sección se denominan fórmulas de *Gauss-Legendre*. Antes de describir el procedimiento, mostraremos que las fórmulas de integración numérica, como la regla del trapecio, pueden obtenerse usando el método de coeficientes indeterminados. Este método se empleará después para desarrollar las fórmulas de Gauss-Legendre.

Figura 22.6
a) Representación gráfica de la regla del trapecio como el área bajo la línea recta que une los puntos extremos fijos. *b*) Se obtiene una mejor estimación de la integral tomando el área bajo la línea recta que pasa por dos puntos intermedios. Estos puntos se ubican en una forma adecuada, de tal manera que se equilibran los errores positivo y negativo.

22.4.1 Método de coeficientes indeterminados

En el capítulo 21 obtuvimos la regla del trapecio integrando un polinomio de interpolación lineal y mediante un razonamiento geométrico. El *método de coeficientes indeterminados* ofrece un tercer procedimiento que también tiene utilidad para encontrar otras técnicas de integración, como la cuadratura de Gauss.

Para ilustrar el procedimiento, la ecuación (22.16) se expresa como

$$I \cong c_0 f(a) + c_1 f(b) \tag{22.17}$$

donde las c = constantes. Ahora observe que la regla del trapecio deberá dar resultados exactos cuando la función que se va a integrar es una constante o una línea recta. Dos ecuaciones simples que representan esos casos son $y = 1$ y $y = x$. Ambas se ilustran en la figura 22.7. Así, las siguientes igualdades se deberán satisfacer:

$$c_0 + c_1 = \int_{-(b-a)/2}^{(b-a)/2} 1\, dx$$

y

$$-c_0\frac{b-a}{2} + c_1\frac{b-a}{2} = \int_{-(b-a)/2}^{(b-a)/2} x\, dx$$

o, evaluando las integrales,

$$c_0 + c_1 = b - a$$

y

$$-c_0\frac{b-a}{2}+c_1\frac{b-a}{2}=0$$

Éstas son dos ecuaciones con dos incógnitas que se resuelven para encontrar

$$c_0=c_1=\frac{b-a}{2}$$

que, al sustituirse en la ecuación (22.17), da

$$I=\frac{b-a}{2}f(a)+\frac{b-a}{2}f(b)$$

que es equivalente a la regla del trapecio.

22.4.2 Desarrollo de la fórmula de Gauss-Legendre de dos puntos

Así como en el caso anterior para la obtención de la regla del trapecio, el objetivo de la cuadratura de Gauss es determinar los coeficientes de una ecuación de la forma

$$I \cong c_0 f(x_0) + c_1 f(x_1) \tag{22.18}$$

donde las c = los coeficientes desconocidos. Sin embargo, a diferencia de la regla del trapecio que utiliza puntos extremos fijos a y b, los argumentos de la función x_0 y x_1 no están fijos en los extremos, sino que son incógnitas (figura 22.8). De esta manera, ahora se tienen cuatro incógnitas que deben evaluarse y, en consecuencia, se requieren cuatro condiciones para determinarlas con exactitud.

Así, como con la regla del trapecio, es posible obtener dos de esas condiciones al suponer que la ecuación (22.18) ajusta con exactitud la integral de una constante y de una función lineal. Después, para tener las otras dos condiciones, sólo se ampliará este razonamiento al suponer que también ajusta la integral de una función parabólica ($y = x^2$) y de una cúbica ($y = x^3$). Al hacerlo, se determinan las cuatro incógnitas y además se obtiene una fórmula de integración lineal de dos puntos que es exacta para cúbicas. Las cuatro ecuaciones que habrá que resolver son:

$$c_0 f(x_0)+c_1 f(x_1)=\int_{-1}^{1} 1\,dx=2 \tag{22.19}$$

$$c_0 f(x_0)+c_1 f(x_1)=\int_{-1}^{1} x\,dx=0 \tag{22.20}$$

$$c_0 f(x_0)+c_1 f(x_1)=\int_{-1}^{1} x^2\,dx=\frac{2}{3} \tag{22.21}$$

$$c_0 f(x_0)+c_1 f(x_1)=\int_{-1}^{1} x^3\,dx=0 \tag{22.22}$$

Las ecuaciones (22.19) a (22.22) pueden resolverse simultáneamente para encontrar

$$c_0=c_1=1$$

$$x_0=-\frac{1}{\sqrt{3}}=-0.5773503\ldots$$

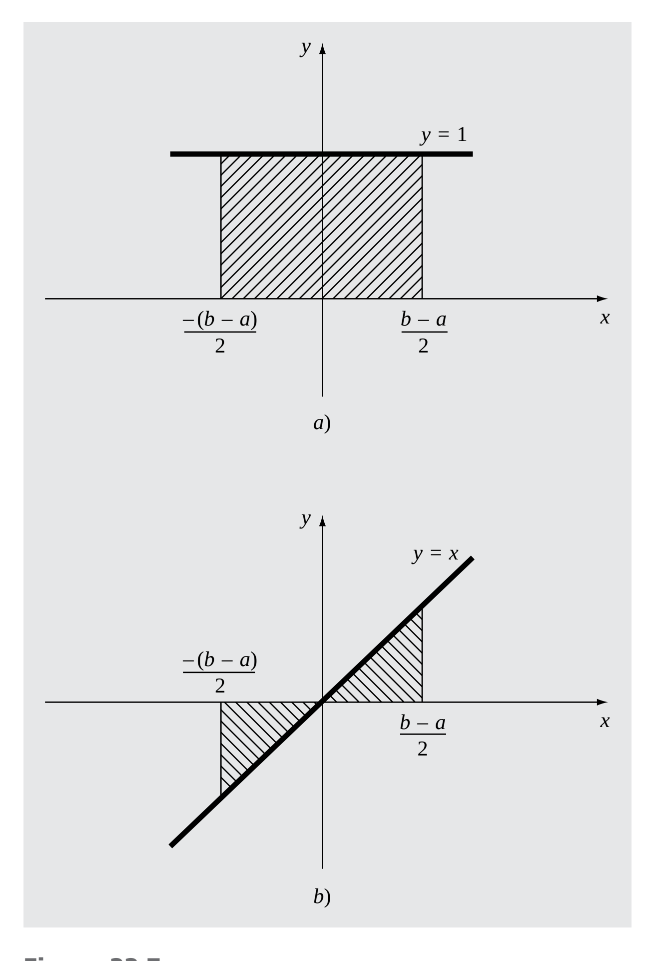


Figura 22.7
Dos integrales que deberán evaluarse exactamente por la regla del trapecio: *a*) una constante y *b*) una línea recta.

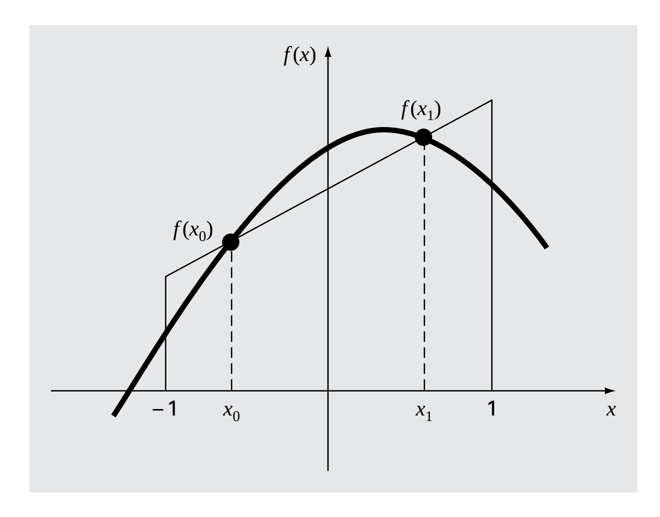


Figura 22.8
Representación gráfica de las variables desconocidas x_0 y x_1 para la integración por medio de la cuadratura de Gauss.

$$x_1 = \frac{1}{\sqrt{3}} = 0.5773503\ldots$$

que se sustituye en la ecuación (22.18) para obtener la fórmula de Gauss-Legendre de dos puntos

$$I \cong f\left(\frac{-1}{\sqrt{3}}\right) + f\left(\frac{1}{\sqrt{3}}\right) \tag{22.23}$$

Así, llegamos al interesante resultado de que la simple suma de los valores de la función en $x = 1/\sqrt{3}$ y $-1/\sqrt{3}$ genera una estimación de la integral que tiene una exactitud de tercer grado.

Observe que los límites de integración en las ecuaciones (22.19) a (22.22) son desde –1 a 1. Esto se hizo para simplificar la matemática y para hacer la formulación tan general como sea posible. Es posible utilizar un simple cambio de variable para transformar otros límites de integración a esta forma. Esto se realiza suponiendo que una nueva variable x_d está relacionada con la variable original x en una forma lineal, como sigue

$$x = a_0 + a_1 x_d \tag{22.24}$$

Si el límite inferior, $x = a$, corresponde a $x_d = -1$, estos valores se sustituyen en la ecuación (22.24):

$$a = a_0 + a_1(-1) \tag{22.25}$$

De manera similar, el límite superior, $x = b$, corresponde a $x_d = 1$, para tener

$$b = a_0 + a_1(1) \tag{22.26}$$

Las ecuaciones (22.25) y (22.26) podrán resolverse simultáneamente para obtener

$$a_0 = \frac{b+a}{2} \tag{22.27}$$

y

$$a_1 = \frac{b-a}{2} \tag{22.28}$$

que se sustituye en la ecuación (22.24) con el siguiente resultado:

$$x = \frac{(b+a)+(b-a)x_d}{2} \tag{22.29}$$

Esta ecuación se diferencia para dar

$$dx = \frac{b-a}{2} dx_d \tag{22.30}$$

Las ecuaciones (22.29) y (22.30) pueden sustituirse ahora por x y dx, respectivamente, en la ecuación que se habrá de integrar. Tales sustituciones efectivamente transforman el intervalo de integración sin cambiar el valor de la integral. El siguiente ejemplo ilustra cómo se hace esto en la práctica.

EJEMPLO 22.3 Fórmula de Gauss-Legendre de dos puntos

Planteamiento del problema Con la ecuación (22.23) evalúe la integral de

$$f(x) = 0.2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$$

entre los límites $x = 0$ y $x = 0.8$. Recuerde que éste fue el problema que resolvimos en el capítulo 21, con las fórmulas de Newton-Cotes. El valor exacto de la integral es 1.640533.

Solución Antes de integrar la función, debemos realizar un cambio de variable para que los límites sean de –1 a +1. Para ello, sustituimos $a = 0$ y $b = 0.8$ en la ecuación (22.29) para obtener

$x = 0.4 + 0.4x_d$

La derivada de esta relación es [ecuación (22.30)]

$dx = 0.4dx_d$

Ambas ecuaciones se sustituyen en la ecuación original para dar

$$\int_0^{0.8} (0.2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5)\, dx$$

$$= \int_{-1}^{1} [0.2 + 25(0.4 + 0.4x_d) - 200(0.4 + 0.4x_d)^2 + 675(0.4 + 0.4x_d)^3$$

$$- 900(0.4 + 0.4x_d)^4 + 400(0.4 + 0.4x_d)^5]0.4dx_d$$

Así, el lado derecho está en la forma adecuada para la evaluación mediante la cuadratura de Gauss. La función transformada se evalúa en $-1/\sqrt{3}$ que es igual a 0.516741 y en $1/\sqrt{3}$ que es igual a 1.305837. Por lo tanto, la integral, de acuerdo con la ecuación (22.23), es

$I \cong 0.516741 + 1.305837 = 1.822578$

que representa un error relativo porcentual de –11.1%. El resultado es comparable en magnitud a la aplicación de la regla del trapecio de cuatro segmentos (tabla 21.1) o a una aplicación simple de las reglas de Simpson 1/3 y 3/8 (ejemplos 21.4 y 21.6). Se espera este último resultado ya que las reglas de Simpson son también de tercer grado de exactitud. Observe que, debido a la elección inteligente de los puntos, la cuadratura de Gauss alcanza esta exactitud considerando tan sólo dos evaluaciones de la función.

22.4.3 Fórmulas con más puntos

Aparte de la fórmula de dos puntos descrita en la sección anterior, se pueden desarrollar versiones con más puntos en la forma general

$$I \cong c_0 f(x_0) + c_1 f(x_1) + \cdots + c_{n-1} f(x_{n-1}) \qquad (22.31)$$

donde n = número de puntos. Los valores de las c y las x para fórmulas de hasta seis puntos se resumen en la tabla 22.1.

EJEMPLO 22.4 Fórmula de Gauss-Legendre de tres puntos

Planteamiento del problema Use la fórmula de tres puntos con la tabla 22.1 para estimar la integral de la misma función que en el ejemplo 22.3.

Solución De acuerdo con la tabla 22.1, la fórmula de tres puntos es

$I = 0.5555556f(-0.7745967) + 0.8888889f(0) + 0.5555556f(0.7745967)$

que es igual a

$I = 0.2813013 + 0.8732444 + 0.4859876 = 1.640533$

que es exacta.

Como la cuadratura de Gauss requiere evaluaciones de la función en puntos irregularmente espaciados dentro del intervalo de integración, no es apropiada para los casos donde la función no se conoce. Es decir, para problemas que tratan con datos tabulados, será necesario interpolar para el argumento dado. Sin embargo, cuando se conoce la función, su eficiencia es de una ventaja decisiva, en particular cuando se deben realizar muchas evaluaciones de la integral.

Tabla 22.1 Factores de ponderación *c* y argumentos de la función *x* usados en las fórmulas de Gauss-Legendre.

Puntos	Factor de ponderación	Argumentos de la función	Error de truncamiento
2	$c_0 = 1.0000000$	$x_0 = -0.577350269$	$\cong f^{(4)}(\xi)$
	$c_1 = 1.0000000$	$x_1 = 0.577350269$	
3	$c_0 = 0.5555556$	$x_0 = -0.774596669$	$\cong f^{(6)}(\xi)$
	$c_1 = 0.8888889$	$x_1 = 0.0$	
	$c_2 = 0.5555556$	$x_2 = 0.774596669$	
4	$c_0 = 0.3478548$	$x_0 = -0.861136312$	$\cong f^{(8)}(\xi)$
	$c_1 = 0.6521452$	$x_1 = -0.339981044$	
	$c_2 = 0.6521452$	$x_2 = 0.339981044$	
	$c_3 = 0.3478548$	$x_3 = 0.861136312$	
5	$c_0 = 0.2369269$	$x_0 = -0.906179846$	$\cong f^{(10)}(\xi)$
	$c_1 = 0.4786287$	$x_1 = -0.538469310$	
	$c_2 = 0.5688889$	$x_2 = 0.0$	
	$c_3 = 0.4786287$	$x_3 = 0.538469310$	
	$c_4 = 0.2369269$	$x_4 = 0.906179846$	
6	$c_0 = 0.1713245$	$x_0 = -0.932469514$	$\cong f^{(12)}(\xi)$
	$c_1 = 0.3607616$	$x_1 = -0.661209386$	
	$c_2 = 0.4679139$	$x_2 = -0.238619186$	
	$c_3 = 0.4679139$	$x_3 = 0.238619186$	
	$c_4 = 0.3607616$	$x_4 = 0.661209386$	
	$c_5 = 0.1713245$	$x_5 = 0.932469514$	

EJEMPLO 22.5 Aplicación de la cuadratura de Gauss al problema del paracaidista en caída

Planteamiento del problema En el ejemplo 21.3 se usó la regla del trapecio de aplicación múltiple para evaluar

$$d = \frac{gm}{c} \int_0^{10} [1 - e^{-(c/m)t}]\, dt$$

donde $g = 9.8$, $c = 12.5$ y $m = 68.1$. El valor exacto de la integral se determinó por medio del cálculo, igual a 289.4351. Recuerde que la mejor estimación obtenida usando la regla del trapecio con 500 segmentos fue 289.4348 con un $|\varepsilon_t| \cong 1.15 \times 10^{-4}\%$. Repita este cálculo usando la cuadratura de Gauss.

Solución Después de modificar la función, se obtienen los siguientes resultados:

Estimación con dos puntos = 290.0145
Estimación con tres puntos = 289.4393
Estimación con cuatro puntos = 289.4352
Estimación con cinco puntos = 289.4351
Estimación con seis puntos = 289.4351

Así, las estimaciones con cinco y seis puntos dan resultados que son exactos hasta la séptima cifra significativa.

22.4.4 Análisis del error en la cuadratura de Gauss

El error en las fórmulas de Gauss-Legendre por lo general se especifica mediante (Carnahan y colaboradores, 1969)

$$E_t = \frac{2^{2n+3}[(n+1)!]^4}{(2n+3)[(2n+2)!]^3} f^{(2n+2)}(\xi) \tag{22.32}$$

donde n = el número de puntos menos uno y $f^{(2n+2)}(\xi)$ = la $(2n + 2)$-ésima derivada de la función, después del cambio de variable con ξ localizada en algún lugar en el intervalo desde –1 hasta 1. Una comparación de la ecuación (22.32) con la tabla 21.2 indica la superioridad de la cuadratura de Gauss respecto a las fórmulas de Newton-Cotes, siempre que las derivadas de orden superior no aumenten sustancialmente cuando se incremente n. El problema 22.8 al final de este capítulo ilustra un caso donde las fórmulas de Gauss-Legendre tienen un desempeño pobre. En tales situaciones, se prefieren la regla de Simpson de aplicación múltiple o la integración de Romberg. No obstante, en muchas de las funciones encontradas en la práctica de la ingeniería, la cuadratura de Gauss proporciona un medio eficiente para la evaluación de las integrales.

22.5 INTEGRALES IMPROPIAS

Hasta aquí, hemos analizado exclusivamente integrales que tienen límites finitos e integrandos acotados. Aunque esos tipos son de uso común en ingeniería, habrá ocasiones en que se deban evaluar integrales impropias. En esta sección nos ocuparemos de un tipo de integral impropia. Es decir, una con límite inferior –∞ y/o límite superior +∞.

Tales integrales a menudo se resuelven con un cambio de variable, que transforma el límite infinito en uno finito. La siguiente identidad sirve para este propósito y trabaja con cualquier función decreciente hacia cero, por lo menos tan rápido como $1/x^2$, conforme x se aproxima a infinito:

$$\int_a^b f(x)\,dx = \int_{1/b}^{1/a} \frac{1}{t^2} f\left(\frac{1}{t}\right) dt \tag{22.33}$$

para $ab > 0$. Por lo tanto, se utiliza sólo cuando a es positiva y b es ∞, o cuando a es –∞ y b es negativa. En los casos donde los límites son desde –∞ a un valor positivo o desde un valor negativo a ∞, la integral puede calcularse en dos partes. Por ejemplo,

$$\int_{-\infty}^{b} f(x)\,dx = \int_{-\infty}^{-A} f(x)\,dx + \int_{-A}^{b} f(x)\,dx \tag{22.34}$$

donde $-A$ se elige como un valor negativo lo suficientemente grande para que la función comience a aproximarse a cero, en forma asintótica por lo menos tan rápido como $1/x^2$. Después que la integral se divide en dos partes, la primera podrá evaluarse con la ecuación (22.33) y la segunda con una fórmula cerrada de Newton-Cotes como la regla de Simpson 1/3.

Un problema al usar la ecuación (22.33) para evaluar una integral es que la función transformada será singular en uno de los límites. Pueden usarse las fórmulas de integración abierta para evitar este dilema, ya que permiten la estimación de la integral sin emplear los puntos extremos del intervalo de integración. Para tener máxima flexibilidad, se requiere una de las fórmulas abiertas vistas en la tabla 21.4 de aplicación múltiple.

Las fórmulas abiertas de aplicación múltiple podrán combinarse con las fórmulas cerradas para los segmentos interiores y fórmulas abiertas para los extremos. Por ejemplo, si se combinan la regla del trapecio de segmentos múltiples y la regla del punto medio se obtiene

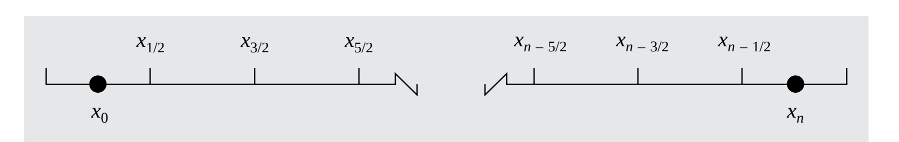


Figura 22.9
Colocación de datos para los límites de integración en la regla extendida del punto medio.

$$\int_{x_0}^{x_n} f(x)\,dx = h\left[\frac{3}{2}f(x_1) + \sum_{i=2}^{n-2} f(x_i) + \frac{3}{2}f(x_{n-1})\right]$$

Además, es posible desarrollar fórmulas semiabiertas para casos donde uno u otro extremo del intervalo es cerrado. Por ejemplo, una fórmula que es abierta en el límite inferior y cerrada en el superior está dada como sigue:

$$\int_{x_0}^{x_n} f(x)\,dx = h\left[\frac{3}{2}f(x_1) + \sum_{i=2}^{n-1} f(x_i) + \frac{1}{2}f(x_n)\right]$$

Aunque se pueden usar estas relaciones, una fórmula preferida es (Press y colaboradores, 1992)

$$\int_{x_0}^{x_n} f(x)\,dx = h[f(x_{1/2}) + f(x_{3/2}) + \cdots + f(x_{n-3/2}) + f(x_{n-1/2})] \tag{22.35}$$

que se conoce como la *regla extendida del punto medio*. Observe que esta fórmula se basa en límites de integración que están $h/2$ después y antes del primer y del último dato respectivamente (figura 22.9).

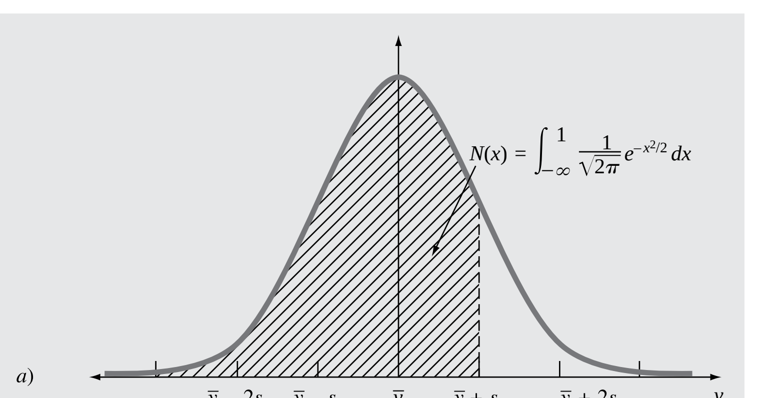
EJEMPLO 22.6 Evaluación de una integral impropia

Planteamiento del problema La *distribución normal acumulativa* es una fórmula importante en estadística (figura 22.10):

$$N(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx \tag{E22.6.1}$$

donde $x = (y - \bar{y})/s_y$ se llama la *desviación estándar normalizada*, la cual representa un cambio de variable para escalar la distribución normal, de tal forma que esté centrada en cero y la distancia a lo largo de la abscisa se mida en múltiplos de la desviación estándar (figura 22.10*b*).

La ecuación (E22.6.1) representa la probabilidad de que un evento sea menor que *x*. Por ejemplo, si *x* = 1, la ecuación (E22.6.1) se utiliza para determinar la probabilidad de que ocurra un evento que es menor que una desviación estándar por arriba de la media, es decir $N(1) = 0.8413$. En otras palabras, si ocurren 100 eventos, aproximadamente 84 serán menores que la media más una desviación estándar. Como la ecuación (E22.6.1) no puede evaluarse de una ma-



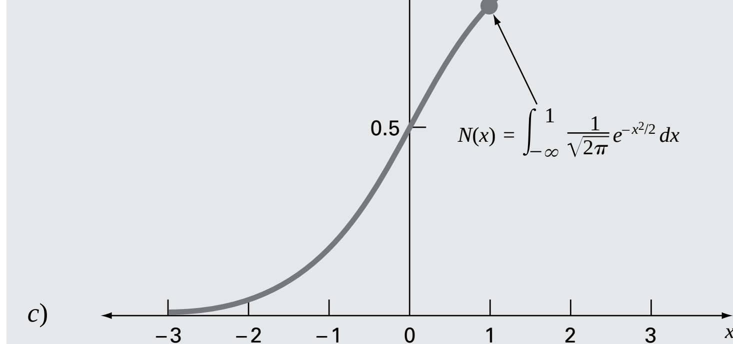


Figura 22.10
a) La distribución normal, *b*) la abscisa transformada en términos de la desviación normal estandarizada y *c*) la distribución normal acumulada. El área sombreada en *a*) y el punto en *c*) representan la probabilidad de que un evento aleatorio sea menor que la media más una desviación estándar.

nera funcional simple, se resuelve numéricamente y se presenta en tablas estadísticas. Con la ecuación (22.34), la regla de Simpson 1/3 y la regla extendida del punto medio determine $N(1)$ en forma numérica.

Solución La ecuación (E22.6.1) se expresa en términos de la ecuación (22.34) como sigue:

$$N(x) = \frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^{-2} e^{-x^2/2}dx + \int_{-2}^{1} e^{-x^2/2}dx\right)$$

La primera integral se evalúa empleando la ecuación (22.33):

$$\int_{-\infty}^{-2} e^{-x^2/2}dx = \int_{-1/2}^{0} \frac{1}{t^2}e^{-1/(2t^2)}dt$$

Después la regla extendida del punto medio con $h = 1/8$ se empleará para estimar

$$\int_{-1/2}^{0} \frac{1}{t^2}e^{-1/(2t^2)}dt \cong \frac{1}{8}[f(x_{-7/16}) + f(x_{-5/16}) + f(x_{-3/16}) + f(x_{-1/16})]$$
$$= \frac{1}{8}[0.3833 + 0.0612 + 0 + 0] = 0.0556$$

Para estimar la segunda integral se usa la regla de Simpson 1/3 con $h = 0.5$, como sigue

$$\int_{-2}^{1} e^{-x^2/2}dx$$
$$= [1 - (-2)]\frac{0.1353 + 4(0.3247 + 0.8825 + 0.8825) + 2(0.6065 + 1) + 0.6065}{3(6)}$$
$$= 2.0523$$

Entonces, el resultado final se calcula mediante

$$N(1) \cong \frac{1}{\sqrt{2\pi}}(0.0556 + 2.0523) = 0.8409$$

que representa un error $\varepsilon_t = 0.046$ por ciento.

El cálculo anterior mejora de diferentes maneras. Primero, se podrían utilizar fórmulas de grado superior; por ejemplo, mediante una integración de Romberg. Segundo, pueden usarse más puntos. Press y colaboradores (2007) exploran con detalle ambas opciones.

Además de los límites infinitos, hay otras formas en las cuales una integral llega a ser impropia. Ejemplos comunes incluyen casos donde la integral es singular tanto en los límites como en un punto dentro de la integral. Press y colaboradores (2007) ofrecen un excelente análisis sobre las formas de manejar esas situaciones.

PROBLEMAS

22.1 Utilice la integración de Romberg de orden h^8 para evaluar

$$\int_0^3 xe^{2x}dx$$

Compare ε_a y ε_t.

22.2 Use la integración de Romberg para evaluar

$$I = \int_1^2 \left(x + \frac{1}{x}\right)^2 dx$$

con una exactitud de ε_s = 0.5% basada en la ecuación (22.9). Debe presentar sus resultados en la forma de la figura 22.3. Utilice la solución analítica de la integral para determinar el error relativo porcentual del resultado obtenido con la integración de Romberg. Verifique que ε_t es menor que el criterio de detención ε_s.

22.3 Emplee la integración de Romberg para evaluar

$$\int_0^2 \frac{e^x \operatorname{sen} x}{1+x^2}\, dx$$

con una exactitud de ε_s = 0.5%. Debe presentar sus resultados en la forma de la figura 22.3.

22.4 Obtenga una estimación de la integral del problema 22.2, pero use fórmulas de Gauss-Legendre con dos, tres y cuatro puntos. Calcule ε_t para cada caso sobre la base de la solución analítica.

22.5 Obtenga una estimación de la integral del problema 22.1, pero use las fórmulas de Gauss-Legendre con dos, tres y cuatro puntos. Calcule ε_t para cada caso sobre la base de la solución analítica.

22.6 Obtenga una estimación de la integral del problema 22.3 con el uso de la fórmula de Gauss-Legendre con cinco puntos.

22.7 Realice el cálculo de los ejemplos 21.3 y 22.5 para el paracaidista que cae, pero use la integración de Romberg (ε_s = 0.05%).

22.8 Emplee fórmulas de Gauss-Legendre de dos a seis puntos para resolver

$$\int_{-3}^3 \frac{1}{1+x^2}\, dx$$

Interprete sus resultados a la luz de la ecuación (22.32).

22.9 Use integración numérica para evaluar lo siguiente:

a) $\displaystyle\int_2^\infty \frac{dx}{x(x+2)}$ *b*) $\displaystyle\int_0^\infty e^{-y}\operatorname{sen}^2 y\, dy$

c) $\displaystyle\int_0^\infty \frac{1}{(1+y^2)(1+y^2/2)}\, dy$ *d*) $\displaystyle\int_{-2}^\infty ye^{-y}dy$

e) $\displaystyle\int_0^\infty \frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx$

Observe que la integral del inciso *e*) es la distribución normal (recuerde la figura 22.10).

22.10 Con base en la figura 22.1, desarrolle un programa de cómputo amigable para el usuario, para segmentos múltiples de las reglas *a*) del trapecio y *b*) de Simpson 1/3. Pruébelo con la integración de

$$\int_0^1 x^{0.1}(1.2-x)(1-e^{20(x-1)})\, dx$$

Utilice el valor verdadero de 0.602298 para calcular ε_t para n = 4.

22.11 Desarrolle un programa de computadora amigable para el usuario para la integración de Romberg, con base en la figura 22.4. Pruébelo con la replicación de los resultados de los ejemplos 22.3 y 22.4, y la función del problema 22.10.

22.12 Desarrolle un programa de computadora amigable para el usuario para cuadratura adaptiva, basado en la figura 22.5. Pruébelo resolviendo el problema 22.10.

22.13 Desarrolle un programa de computadora amigable para el usuario para la cuadratura de Gauss. Pruébelo con la duplicación de los resultados de los ejemplos 22.3 y 22.4, y la función del problema 22.10.

22.14 No existe forma cerrada para la solución de la función de error,

$$\operatorname{erf}(a) = \frac{2}{\sqrt{\pi}}\int_0^a e^{-x^2}dx$$

Emplee el enfoque de la cuadratura de Gauss de dos puntos para estimar erf(1.5). Observe que el valor exacto es 0.966105.

22.15 La cantidad de masa transportada por un tubo durante cierto periodo se calcula con

$$M = \int_{t_1}^{t_2} Q(t)c(t)dt$$

donde M = masa (mg), t_1 = tiempo inicial (min), t_2 = tiempo final (min), $Q(t)$ = tasa de flujo (m^3/min) y $c(t)$ = concentración (mg/m^3). Las representaciones funcionales siguientes definen las variaciones temporales en el flujo y la concentración:

$$Q(t) = 9 + 5\cos^2(0.4t)$$
$$c(t) = 5e^{-0.5t} + 2e^{0.15t}$$

Determine la masa transportada entre t_1 = 2 min y t_2 = 8 min, con integración de Romberg para una tolerancia de 0.1%.

22.16 Las profundidades de un río H se miden a distancias espaciadas iguales a través de un canal como se muestra en la tabla siguiente. El área de la sección transversal del río se determina por integración con

$$A_c = \int_0^x H(x)dx$$

Emplee integración de Romberg para llevar a cabo la integración con un criterio de detención de 1%.

x, m	0	2	4	6	8	10	12	14	16
H, m	0	1.9	2	2	2.4	2.6	2.25	1.12	0

22.17 Recuerde que la velocidad de un paracaidista en caída libre con arrastre lineal se puede calcular analíticamente como

$$v(t) = \frac{gm}{c}(1 - e^{-(c/m)t})$$

donde $v(t)$ = velocidad (m/s), t = tiempo (s), g = 9.81 m/s^2, m = masa (kg), c = coeficiente de arrastre lineal (kg/s). Use la integración de Romberg para calcular la distancia que viaja el paracaidista durante los primeros 8 segundos de caída libre dados m = 80 kg y c = 10 kg/s. Calcule la respuesta hasta ε_s = 1%.

22.18 Pruebe que la ecuación (22.15) es equivalente a la regla de Boole.

CAPÍTULO 23

Diferenciación numérica

En el capítulo 4 ya se introdujo la noción de diferenciación numérica. Recuerde que se emplearon las expansiones en serie de Taylor para obtener las aproximaciones de las derivadas en diferencias finitas divididas. En el mismo capítulo se desarrollaron las aproximaciones en diferencias divididas hacia adelante, hacia atrás y centradas para la primer derivada y las derivadas de orden superior. Recuerde que, en el mejor de los casos, dichas estimaciones tenían errores que fueron $O(h^2)$; es decir, sus errores eran proporcionales al cuadrado del tamaño de paso. Este nivel de exactitud se debe al número de términos de la serie de Taylor que se utilizaron durante la deducción de esas fórmulas. Ahora se mostrará cómo desarrollar fórmulas de mayor exactitud utilizando más términos.

23.1 FÓRMULAS DE DIFERENCIACIÓN CON ALTA EXACTITUD

Como se indicó antes, se pueden generar fórmulas por diferencias divididas de alta exactitud tomando términos adicionales en la expansión de la serie de Taylor. Por ejemplo, la expansión de la serie de Taylor hacia adelante se escribe como [ecuación (4.21)]:

$$f(x_{i+1}) = f(x_i) + f'(x_i)h + \frac{f''(x_i)}{2}h^2 + \cdots \tag{23.1}$$

de la que se despeja

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_i)}{h} - \frac{f''(x_i)}{2}h + O(h^2) \tag{23.2}$$

En el capítulo 4 truncamos este resultado al excluir los términos de la segunda derivada en adelante y nos quedamos con un resultado final,

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_i)}{h} + O(h) \tag{23.3}$$

Ahora retendremos, en cambio, el término de la segunda derivada sustituyendo la siguiente aproximación de la segunda derivada [recuerde la ecuación (4.24)]

$$f''(x_i) = \frac{f(x_{i+2}) - 2f(x_{i+1}) + f(x_i)}{h^2} + O(h) \tag{23.4}$$

en la ecuación (23.2) para dar

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_i)}{h} - \frac{f(x_{i+2}) - 2f(x_{i+1}) + f(x_i)}{2h^2}h + O(h^2)$$

o, al agrupar términos,

$$f'(x_i) = \frac{-f(x_{i+2}) + 4f(x_{i+1}) - 3f(x_i)}{2h} + O(h^2) \tag{23.5}$$

Observe que al incluir el término de la segunda derivada mejora la exactitud a $O(h^2)$. Es posible desarrollar versiones similares mejoradas para las fórmulas hacia atrás y centradas, así como para las aproximaciones de derivadas de orden superior. Las fórmulas se resumen en las figuras 23.1 a 23.3, junto con todos los resultados del capítulo 4. El siguiente ejemplo ilustra la utilidad de esas fórmulas para la estimación de las derivadas.

EJEMPLO 23.1 Fórmulas de diferenciación con alta exactitud

Planteamiento del problema Recuerde que en el ejemplo 4.4 estimamos la derivada de

$$f(x) = -0.1x^4 - 0.15x^3 - 0.5x^2 - 0.25x + 1.2$$

en $x = 0.5$ usando diferencias divididas finitas y un tamaño de paso: $h = 0.25$,

	Hacia adelante $O(h)$	Hacia atrás $O(h)$	Centrada $O(h^2)$
Estimación	–1.155	–0.714	–0.934
ε_t (%)	–26.5	21.7	–2.4

Primera derivada — Error

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_i)}{h} \qquad O(h)$$

$$f'(x_i) = \frac{-f(x_{i+2}) + 4f(x_{i+1}) - 3f(x_i)}{2h} \qquad O(h^2)$$

Segunda derivada

$$f''(x_i) = \frac{f(x_{i+2}) - 2f(x_{i+1}) + f(x_i)}{h^2} \qquad O(h)$$

$$f''(x_i) = \frac{-f(x_{i+3}) + 4f(x_{i+2}) - 5f(x_{i+1}) + 2f(x_i)}{h^2} \qquad O(h^2)$$

Tercera derivada

$$f'''(x_i) = \frac{f(x_{i+3}) - 3f(x_{i+2}) + 3f(x_{i+1}) - f(x_i)}{h^3} \qquad O(h)$$

$$f'''(x_i) = \frac{-3f(x_{i+4}) + 14f(x_{i+3}) - 24f(x_{i+2}) + 18f(x_{i+1}) - 5f(x_i)}{2h^3} \qquad O(h^2)$$

Cuarta derivada

$$f''''(x_i) = \frac{f(x_{i+4}) - 4f(x_{i+3}) + 6f(x_{i+2}) - 4f(x_{i+1}) + f(x_i)}{h^4} \qquad O(h)$$

$$f''''(x_i) = \frac{-2f(x_{i+5}) + 11f(x_{i+4}) - 24f(x_{i+3}) + 26f(x_{i+2}) - 14f(x_{i+1}) + 3f(x_i)}{h^4} \qquad O(h^2)$$

Figura 23.1
Fórmulas de diferencias finitas divididas hacia adelante: se presentan dos versiones para cada derivada. La última versión emplea más términos de la expansión de la serie de Taylor y, en consecuencia, es más exacta.

donde los errores fueron calculados basándose en el valor verdadero: –0.9125. Repita este cálculo, pero ahora emplee las fórmulas con alta exactitud a partir de las figuras 23.1 a 23.3.

Solución Los datos necesarios para este ejemplo son

$x_{i-2} = 0$ $\qquad f(x_{i-2}) = 1.2$

$x_{i-1} = 0.25$ $\qquad f(x_{i-1}) = 1.1035156$

$x_i = 0.5$ $\qquad f(x_i) = 0.925$

Primera derivada | Error

$$f'(x_i) = \frac{f(x_i) - f(x_{i-1})}{h} \qquad O(h)$$

$$f'(x_i) = \frac{3f(x_i) - 4f(x_{i-1}) + f(x_{i-2})}{2h} \qquad O(h^2)$$

Segunda derivada

$$f''(x_i) = \frac{f(x_i) - 2f(x_{i-1}) + f(x_{i-2})}{h^2} \qquad O(h)$$

$$f''(x_i) = \frac{2f(x_i) - 5f(x_{i-1}) + 4f(x_{i-2}) - f(x_{i-3})}{h^2} \qquad O(h^2)$$

Tercera derivada

$$f'''(x_i) = \frac{f(x_i) - 3f(x_{i-1}) + 3f(x_{i-2}) - f(x_{i-3})}{h^3} \qquad O(h)$$

$$f'''(x_i) = \frac{5f(x_i) - 18f(x_{i-1}) + 24f(x_{i-2}) - 14f(x_{i-3}) + 3f(x_{i-4})}{2h^3} \qquad O(h^2)$$

Cuarta derivada

$$f''''(x_i) = \frac{f(x_i) - 4f(x_{i-1}) + 6f(x_{i-2}) - 4f(x_{i-3}) + f(x_{i-4})}{h^4} \qquad O(h)$$

$$f''''(x_i) = \frac{3f(x_i) - 14f(x_{i-1}) + 26f(x_{i-2}) - 24f(x_{i-3}) + 11f(x_{i-4}) - 2f(x_{i-5})}{h^4} \qquad O(h^2)$$

Figura 23.2
Fórmulas de diferencias finitas divididas hacia atrás: se presentan dos versiones para cada derivada. La última versión emplea más términos de la expansión de la serie de Taylor y, en consecuencia, es más exacta.

Primera derivada | Error

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_{i-1})}{2h} \qquad O(h^2)$$

$$f'(x_i) = \frac{-f(x_{i+2}) + 8f(x_{i+1}) - 8f(x_{i-1}) + f(x_{i-2})}{12h} \qquad O(h^4)$$

Segunda derivada

$$f''(x_i) = \frac{f(x_{i+1}) - 2f(x_i) + f(x_{i-1})}{h^2} \qquad O(h^2)$$

$$f''(x_i) = \frac{-f(x_{i+2}) + 16f(x_{i+1}) - 30f(x_i) + 16f(x_{i-1}) - f(x_{i-2})}{12h^2} \qquad O(h^4)$$

Tercera derivada

$$f'''(x_i) = \frac{f(x_{i+2}) - 2f(x_{i+1}) + 2f(x_{i-1}) - f(x_{i-2})}{2h^3} \qquad O(h^2)$$

$$f'''(x_i) = \frac{-f(x_{i+3}) + 8f(x_{i+2}) - 13f(x_{i+1}) + 13f(x_{i-1}) - 8f(x_{i-2}) + f(x_{i-3})}{8h^3} \qquad O(h^4)$$

Cuarta derivada

$$f''''(x_i) = \frac{f(x_{i+2}) - 4f(x_{i+1}) + 6f(x_i) - 4f(x_{i-1}) + f(x_{i-2})}{h^4} \qquad O(h^2)$$

$$f''''(x_i) = \frac{-f(x_{i+3}) + 12f(x_{i+2}) + 39f(x_{i+1}) + 56f(x_i) - 39f(x_{i-1}) + 12f(x_{i-2}) + f(x_{i-3})}{6h^4} \qquad O(h^4)$$

Figura 23.3
Fórmulas de diferencias finitas divididas centradas: se presentan dos versiones para cada derivada. La última versión emplea más términos de la expansión de la serie de Taylor y, en consecuencia, es más exacta.

$x_{i+1} = 0.75$ $\quad f(x_{i+1}) = 0.6363281$

$x_{i+2} = 1$ $\quad f(x_{i+2}) = 0.2$

La diferencia hacia adelante de exactitud $O(h^2)$ se calcula como sigue (figura 23.1):

$$f'(0.5) = \frac{-0.2 + 4(0.6363281) - 3(0.925)}{2(0.25)} = -0.859375 \qquad \varepsilon_t = 5.82\%$$

La diferencia hacia atrás de exactitud $O(h^2)$ se calcula como (figura 23.2):

$$f'(0.5) = \frac{3(0.925) - 4(1.1035156) + 1.2}{2(0.25)} = -0.878125 \qquad \varepsilon_t = 3.77\%$$

La diferencia centrada de exactitud $O(h^4)$ se calcula como (figura 23.3):

$$f'(0.5) = \frac{-0.2 + 8(0.6363281) - 8(1.1035156) + 1.2}{12(0.25)} = -0.9125 \qquad \varepsilon_t = 0\%$$

Como se esperaba, los errores para las diferencias hacia adelante y hacia atrás son considerablemente menores y los resultados más exactos que los del ejemplo 4.4. Sin embargo, de manera sorprendente, la diferencia centrada da un resultado perfecto. Esto es porque las fórmulas se basan en la serie de Taylor, y son equivalentes a polinomios que pasan a través de los puntos asociados con los datos.

23.2 EXTRAPOLACIÓN DE RICHARDSON

Hasta aquí hemos visto que hay dos formas para mejorar la estimación obtenida al emplear diferencias finitas divididas: 1) disminuir el tamaño de paso o 2) usar una fórmula de grado superior que emplee más puntos. Un tercer procedimiento, basado en la extrapolación de Richardson, utiliza dos estimaciones de la derivada para calcular una tercera aproximación más exacta.

Recuerde de la sección 22.2.1 que la extrapolación de Richardson constituye un medio para obtener una mejor estimación de la integral I por medio de la fórmula [ecuación (22.4)]

$$I \cong I(h_2) + \frac{1}{(h_1/h_2)^2 - 1}[I(h_2) - I(h_1)] \tag{23.6}$$

donde $I(h_1)$ e $I(h_2)$ son estimaciones de la integral obtenidas usando dos tamaños de paso, h_1 y h_2. Debido a su conveniencia cuando se expresa como un algoritmo computacional, esta fórmula usualmente se escribe para el caso en que $h_2 = h_1/2$, como

$$I \cong \frac{4}{3}I(h_2) - \frac{1}{3}I(h_1) \tag{23.7}$$

De manera similar, la ecuación (23.7) se escribirá para las derivadas como

$$D \cong \frac{4}{3}D(h_2) - \frac{1}{3}D(h_1) \tag{23.8}$$

Para aproximaciones por diferencia centrada con $O(h^2)$, la aplicación de esta fórmula dará una nueva estimación de la derivada de $O(h^4)$.

EJEMPLO 23.2 Extrapolación de Richardson

Planteamiento del problema Utilizando la misma función que en el ejemplo 23.1, estime la primera derivada en $x = 0.5$ empleando tamaños de paso $h_1 = 0.5$ y $h_2 = 0.25$. Después, con la ecuación (23.8) calcule una mejor estimación con la extrapolación de Richardson. Recuerde que el valor verdadero es –0.9125.

Solución Las estimaciones de la primera derivada se calculan con diferencias centradas como sigue:

$$D(0.5) = \frac{0.2 - 1.2}{1} = -1.0 \quad \varepsilon_t = -9.6\%$$

y

$$D(0.25) = \frac{0.6363281 - 1.1035156}{0.5} = -0.934375 \quad \varepsilon_t = -2.4\%$$

Se determina una mejor estimación aplicando la ecuación (23.8) al obtener

$$D = \frac{4}{3}(-0.934375) - \frac{1}{3}(-1) = -0.9125$$

que, en este caso, es un resultado perfecto.

El ejemplo anterior dio un resultado perfecto debido a que la función analizada era un polinomio de cuarto grado. El resultado perfecto se debió al hecho de que la extrapolación de Richardson, en realidad, es equivalente a ajustar un polinomio de grado superior a los datos y después evaluar las derivadas con diferencias divididas centradas. Así, este caso concuerda, precisamente, con la derivada del polinomio de cuarto grado. Para las otras funciones que no son polinomios, por supuesto, esto no ocurrirá y nuestra estimación de la derivada será mejor, pero no perfecta. En consecuencia, como en el caso de la aplicación de la extrapolación de Richardson, el procedimiento puede aplicarse de manera iterativa usando un algoritmo de Romberg, hasta que el resultado se halle por debajo de un criterio de error aceptable.

23.3 DERIVADAS DE DATOS IRREGULARMENTE ESPACIADOS

Los procedimientos analizados hasta ahora se diseñaron principalmente para determinar la derivada de una función dada. Para las aproximaciones por diferencias finitas divididas de la sección 23.1, los datos deben estar igualmente espaciados. Para la técnica de extrapolación de Richardson de la sección 23.2, los datos deben estar igualmente espaciados y generados por sucesivas divisiones a la mitad de los intervalos. Para tener un buen control del espaciamiento de datos, con frecuencia, sólo es posible cuando se utiliza una función para generar la tabla de valores.

Sin embargo, la información empírica (es decir, datos a partir de experimentos o de estudios de campo) con frecuencia se obtiene a intervalos desiguales. Tal información no puede analizarse con las técnicas estudiadas hasta aquí.

Una manera de emplear datos irregularmente espaciados consiste en ajustar un polinomio de interpolación de Lagrange de segundo grado [recuerde la ecuación (18.23)] a cada conjunto de tres puntos adyacentes. Recuerde que estos polinomios no requieren que los puntos estén igualmente espaciados. Si se deriva analíticamente el polinomio de segundo grado se obtiene

$$f'(x) = f(x_{i-1})\frac{2x - x_i - x_{i+1}}{(x_{i-1} - x_i)(x_{i-1} - x_{i+1})} + f(x_i)\frac{2x - x_{i-1} - x_{i+1}}{(x_i - x_{i-1})(x_i - x_{i+1})} + f(x_{i+1})\frac{2x - x_{i-1} - x_i}{(x_{i+1} - x_{i-1})(x_{i+1} - x_i)} \quad (23.9)$$

donde x es el valor en el cual se quiere estimar la derivada. Aunque esta ecuación es más complicada que las aproximaciones de la primera derivada de las figuras 23.1 a 23.3, tiene importantes ventajas. Primero, sirve para estimar la derivada en cualquier punto dentro de un intervalo determinado por los tres puntos. Segundo, los puntos no tienen que estar igualmente espaciados y tercero, la estimación de la derivada tiene la misma exactitud que la diferencia centrada [ecuación (4.22)]. De hecho, con puntos igualmente espaciados, la ecuación (23.9) evaluada en $x = x_i$ se reduce a la ecuación (4.22).

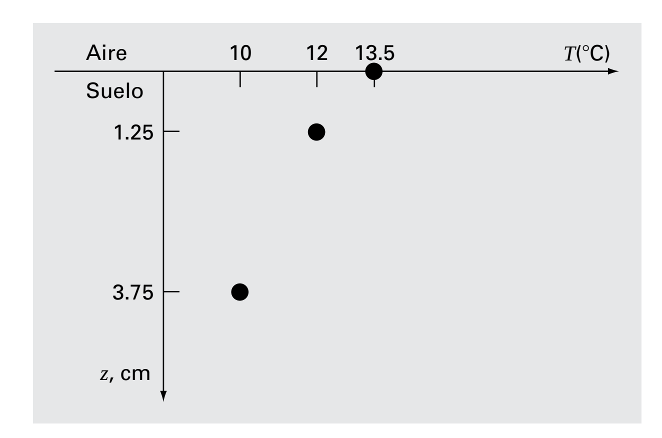


Figura 23.4
Temperatura contra la profundidad bajo el suelo.

EJEMPLO 23.3 Diferenciación de datos irregularmente espaciados

Planteamiento del problema Como se muestra en la figura 23.4, un gradiente de temperatura puede medirse abajo del suelo. El flujo de calor en la interfaz suelo-aire puede calcularse mediante la ley de Fourier,

$$q(z=0) = -k\rho C \left.\frac{dT}{dz}\right|_{z=0}$$

donde q = flujo de calor (W/m^2), k = coeficiente de difusividad térmica en el suelo ($\cong 3.5 \times 10^{-7}$ m^2/s), ρ = densidad del suelo ($\cong$ 1 800 kg/m^3) y C = calor específico del suelo ($\cong$ 840 J/(kg · °C)). Observe que un valor positivo del flujo indica que el calor se transfiere del aire al suelo. Utilice diferenciación numérica para evaluar el gradiente en la interfaz suelo-aire y emplee dicha estimación para determinar el flujo de calor bajo el suelo.

Solución La ecuación (23.9) se utiliza para calcular la derivada como sigue

$$f'(x) = 13.5\frac{2(0)-1.25-3.75}{(0-1.25)(0-3.75)} + 12\frac{2(0)-0-3.75}{(1.25-0)(1.25-3.75)}$$
$$+10\frac{2(0)-0-1.25}{(3.75-0)(3.75-1.25)}$$
$$= -14.4 + 14.4 - 1.333333 = -1.333333°\text{C/cm}$$

que al ser sustituida se obtiene (advierta que 1 W = 1 J/s),

$$q(z=0) = -3.5 \times 10^{-7}\,\frac{\text{m}^2}{\text{s}}\left(1\,800\,\frac{\text{kg}}{\text{m}^3}\right)\left(840\,\frac{\text{J}}{\text{kg}\cdot°\text{C}}\right)\left(-133.3333\,\frac{°\text{C}}{\text{m}}\right)$$
$$= 70.56 \text{ W/m}^2$$

23.4 DERIVADAS E INTEGRALES PARA DATOS CON ERRORES

Además de tener espaciados irregulares, otro problema en la diferenciación de datos empíricos es que generalmente se presentan errores de medición. Una desventaja de la diferenciación numérica es que tiende a amplificar los errores de los datos. La figura 23.5*a* muestra datos uniformes sin errores, que al diferenciarse numéricamente producen un resultado adecuado (figura 23.5*c*). En cambio, la figura 23.5*b* usa los mismos datos, pero con algunos puntos ligeramente por arriba y otros por abajo. Esta pequeña modificación es apenas notoria en la figura 23.5*b*. Sin embargo, el efecto resultante en la figura 23.5*d* es significativo, ya que el proceso de diferenciación amplifica los errores.

Como era de esperarse, el principal procedimiento para determinar derivadas de datos imprecisos consiste en usar regresión por mínimos cuadrados, para ajustar una función suave y diferenciable a los datos. Si no se tiene alguna otra información, una regresión polinomial de grado inferior podría ser una buena elección. Obviamente, si se conoce la verdadera relación funcional entre las variables dependiente e independiente, esta relación deberá ser la base para el ajuste por mínimos cuadrados.

23.4.1 Diferenciación *versus* integración de datos inciertos

Así como las técnicas para el ajuste de curvas, como la regresión, se utilizan para diferenciar datos inciertos, se emplea un proceso similar para la integración. No obstante, debido a la diferencia en estabilidad entre diferenciación e integración, esto rara vez se hace.

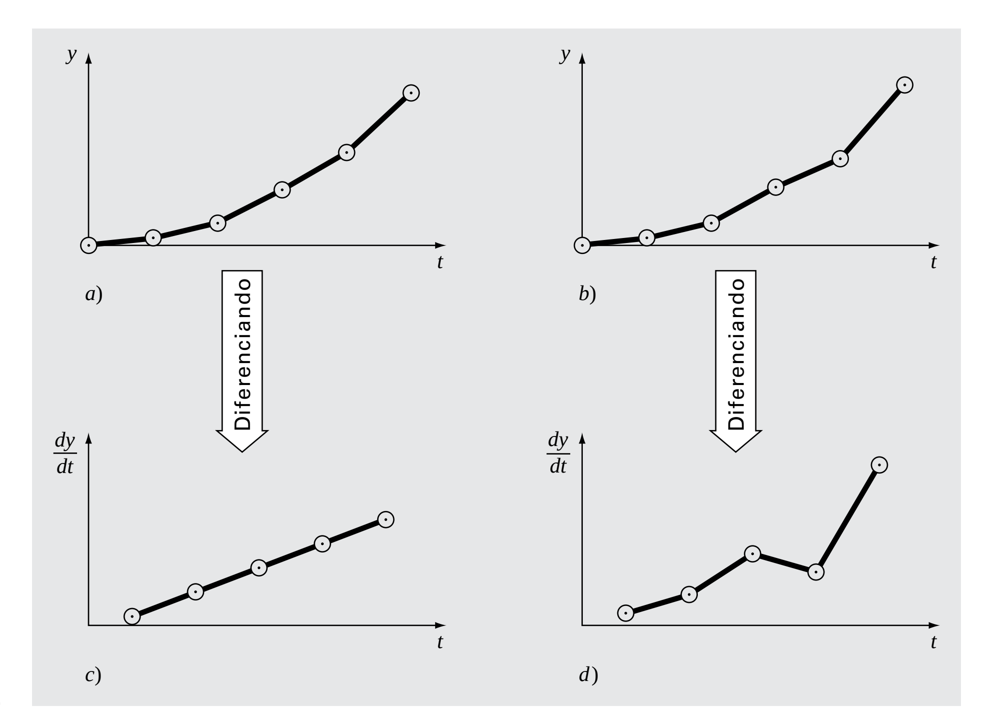


Figura 23.5
Ilustración de cómo los pequeños errores en los datos se amplifican mediante la diferenciación numérica: *a*) datos sin error, *b*) datos modificados ligeramente, *c*) resultado de la diferenciación numérica que se obtiene de la curva *a*), y *d*) la diferenciación resultante de la curva *b*) que manifiesta un aumento en la variabilidad. En cambio, la operación inversa de integración [moviéndose de *d*) a *b*) y tomando el área bajo *d*)] tienden a suavizar o atenuar los errores en los datos.

Como se ilustra en la figura 23.5, la diferenciación tiende a ser inestable; es decir, amplifica los errores. En cambio, el hecho de que la integración sea un proceso de suma tiende a hacerlo muy estable respecto a datos inciertos. En esencia, conforme los puntos se suman para formar una integral, los errores aleatorios positivos y negativos tienden a compensarse. En cambio, debido a que la diferenciación es sustractiva, los errores aleatorios positivos y negativos tienden a sumarse.

23.5 DERIVADAS PARCIALES

Las derivadas parciales a lo largo de una sola dimensión se calculan del mismo modo que las derivadas ordinarias. Por ejemplo, suponga que se quiere determinar las derivadas parciales para una función bidimensional, $f(x, y)$. Para datos igualmente espaciados, las primeras derivadas parciales se pueden aproximar con diferencias centradas,

$$\frac{\partial f}{\partial x} = \frac{f(x+\Delta x, y) - f(x-\Delta x, y)}{2\Delta x} \tag{23.10}$$

$$\frac{\partial f}{\partial y} = \frac{f(x, y+\Delta y) - f(x, y-\Delta y)}{2\Delta y} \tag{23.11}$$

Todas las demás fórmulas y todos los demás enfoques que se han expuesto hasta ahora se pueden aplicar para evaluar derivadas parciales de un modo similar.

Para derivadas de mayor orden, se podría desear diferenciar una función con respecto a dos o más variables diferentes. El resultado se llama *derivada parcial mixta*. Por ejemplo, se puede querer tomar la derivada parcial de $f(x, y)$ con respecto a ambas variables independientes

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) \tag{23.12}$$

Para desarrollar una aproximación con diferencia finita, se puede primero formar una diferencia en x de las derivadas parciales en y,

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\dfrac{\partial f}{\partial y}(x+\Delta x, y) - \dfrac{\partial f}{\partial y}(x-\Delta x, y)}{2\Delta x} \tag{23.13}$$

Luego se pueden usar las diferencias finitas para evaluar cada una de las parciales en *y*,

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\frac{f(x+\Delta x, y+\Delta y) - f(x+\Delta x, y-\Delta y)}{2\Delta y} - \frac{f(x-\Delta x, y+\Delta y) - f(x-\Delta x, y-\Delta y)}{2\Delta y}}{2\Delta x} \tag{23.14}$$

Agrupando términos se obtiene el resultado final

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{f(x+\Delta x, y+\Delta y) - f(x+\Delta x, y-\Delta y) - f(x-\Delta x, y+\Delta y) + f(x-\Delta x, y-\Delta y)}{4\Delta x \Delta y} \tag{23.15}$$

SOFTWARE

23.6 INTEGRACIÓN/DIFERENCIACIÓN NUMÉRICAS CON PAQUETES DE SOFTWARE

Los paquetes de software tienen muchas capacidades para la integración y la diferenciación numérica. En esta sección le daremos una muestra de algunas de las más útiles.

23.6.1 MATLAB

MATLAB tiene una variedad de funciones incorporadas que permiten integrar y diferenciar funciones y datos (tabla 23.1). En esta sección se ilustrarán algunas de estas capacidades.

Tabla 23.1 Funciones MATLAB para implementar *a*) integración y *b*) diferenciación.

Función	Descripción
***a*) Integración:**	
cumtrapz	Integración trapezoidal numérica acumulativa
dblquad	Evaluar numéricamente integral doble
polyint	Integrar polinomio analíticamente
quad	Evaluar numéricamente integral, cuadratura adaptiva Simpson
quadgk	Evaluar numéricamente integral, cuadratura adaptiva Gauss-Kronrod
quadl	Evaluación numérica integral, cuadratura adaptiva Lobatto
quadv	Cuadratura vectorizada
trapz	Integración trapezoidal numérica
triplequad	Evaluar numéricamente integral triple
***b*) Diferenciación:**	
del2	Laplaciana discreta
diff	Diferencias y derivadas aproximadas
gradient	Gradiente numérico
polyder	Derivada de polinomio

MATLAB puede integrar tanto datos discretos como funciones. Por ejemplo, `trapz` calcula la integral de valores discretos usando la regla trapezoidal de aplicaciones múltiples. La siguiente es una sencilla representación de su sintaxis:

```
q = trapz(x, y)
```

donde los dos vectores, `x` y `y`, contienen las variables independiente y dependiente, respectivamente, y `q` contiene la integral resultante. También tiene otra función, `cumtrapz` que calcula la integral acumulativa. Para este caso, el resultado es un vector cuyos elementos `q(k)` contienen la integral de `x (1)` a `x (k)`.

Cuando el integrando está disponible en forma funcional, `quad` genera la integral definida usando cuadratura adaptiva. La siguiente es una representación sencilla de su sintaxis:

```
q = quad(fun, a, b)
```

donde `fun` es la función que se va a integrar y `a` y `b` son los límites de integración.

EJEMPLO 23.4 Uso de la integración numérica para calcular distancia a partir de velocidad

Planteamiento del problema Como se describió en la sección PT6.1, se puede usar integración para calcular la distancia, $y(t)$, de un objeto con base en su velocidad, $v(t)$, como en

$$y(t) = \int_0^t v(t)\,dt \tag{E23.4.1}$$

Recuerde de la sección 1.1 que la velocidad de un paracaidista en caída libre, sujeto a arrastre lineal y con velocidad inicial cero, se puede calcular con

$$v(t) = \frac{gm}{c}\left(1 - e^{-\frac{c}{m}t}\right) \tag{E23.4.2}$$

Si se sustituye la ecuación (E23.4.2) en la ecuación (E23.4.1), el resultado se puede integrar analíticamente, con la condición inicial, $y(0) = 0$, para dar

$$y(t) = \frac{gm}{c}t - \frac{gm^2}{c^2}\left(1 - e^{-\frac{c}{m}t}\right)$$

Este resultado se puede usar para calcular que un paracaidista de 70 kg con un coeficiente de arrastre de 12.5 kg/s caerá 799.73 m durante un periodo de 20 s.

Use funciones MATLAB para realizar numéricamente la misma integración. Además, desarrolle una gráfica de las distancias analíticas y calculadas numéricamente, junto con la velocidad, en la misma gráfica.

Solución Se puede usar primero la ecuación (E23.4.2) para generar algunos tiempos y algunas velocidades con espaciamiento desigual. Se puede entonces redondear estas velocidades de modo que se parezcan más a los valores medidos; es decir, no son exactas,

```
>> format short g
>> t=[0 1 2 3 4.3 7 12 16];
>> g=9.81;m=70;c=12.5;
>> v=round(g*m/c*(1-exp(-c/m*t)));
```

La distancia total se puede entonces calcular como

```
>> y=trapz(t,v)

y =
  789.6
```

Así, después de 20 segundos, el paracaidista ha caído 789.6 m, que es razonablemente cercano a la solución exacta analítica de 799.73 m.

Si se quiere tener la distancia recorrida acumulativa a cada momento, se puede usar `cumtrapz` para calcular

```
>> yc=cumtrapz(t,v)

yc =
  0  4.5  17  36.5  70.3  162.1  379.6  579.6  789.6
```

Se genera una gráfica de soluciones numéricas y analíticas junto con velocidades exactas y redondeadas, con los siguientes comandos:

```
>> ta=linspace(t(1),t(length(t)));
>> ya=g*m/c*ta-g*m^2/c^2*(1-exp(-c/m*ta));
>> plot(ta,ya,t,yc,'o')
>> title('Distance versus time')
>> xlabel('t (s)'),ylabel('x (m)')
>> legend('analytical','numerical')
```

Como en la figura 23.6, los resultados numéricos y analíticos coinciden bastante bien.

SOFTWARE

Finalmente, la función `quad` se puede usar para evaluar la integral con cuadratura adaptiva

```
>> va=@(t) g*m/c*(1-exp(-c/m*t));
>> yq=quad(va,t(1),t(length(t)))

yq =
  799.73
```

Este resultado es idéntico a la solución analítica hasta dentro de los 5 dígitos que se muestran en pantalla.

Como se lista en la tabla 23.1*b*, MATLAB tiene diversas funciones incorporadas para evaluar derivadas, incluidas las funciones `diff` y `gradient`. Cuando se le pasa un vector unidimensional de longitud *n*, la función `diff` devuelve un vector de longitud $n - 1$ que contiene las diferencias entre elementos adyacentes. Esto se puede entonces usar para determinar aproximaciones con diferencia finita de primeras derivadas.

La función `gradient` también devuelve diferencias. Sin embargo, lo hace de un modo más compatible con la evaluación de derivadas en los valores mismos que en los intervalos entre valores. La siguiente es una representación sencilla de su sintaxis:

```
fx = gradient(f)
```

donde `f` = un vector unidimensional de longitud *n*, y `fx` es un vector de longitud *n* que contiene las diferencias con base en `f`. Igual que con la función `diff`, el primer valor devuelto es la diferencia entre el primer valor y el segundo. Sin embargo, para los valores intermedios, se devuelve una diferencia centrada basada en los valores adyacentes,

$$diff_i = \frac{f_{i+1} - f_{i-1}}{2}$$

Entonces se calcula el último valor como la diferencia entre los dos valores finales. Por lo tanto, los resultados equivalen a usar diferencias centradas para todos los valores intermedios con diferencias hacia adelante y hacia atrás en los extremos.

Observe que se supone que el espaciamiento entre puntos es uno. Si el vector representa datos igualmente espaciados, la siguiente versión divide todos los resultados entre el intervalo y, por lo tanto, devuelve los valores reales de las derivadas,

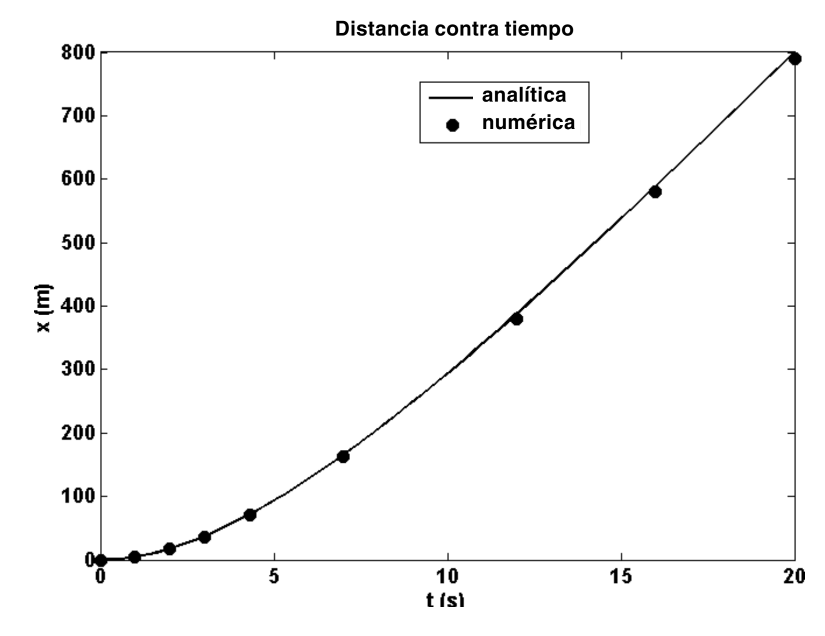


Figura 23.6
Gráfica de distancia contra tiempo. La línea se calculó con la solución analítica, mientras que los puntos se determinaron numéricamente con la función `cumtrapz`.

```
fx = gradient(f, h)
```

donde *h* = espaciamiento entre puntos.

EJEMPLO 23.5 Uso de `diff` y `gradient` para diferenciación

Planteamiento del problema Explore cómo se pueden usar las funciones `diff` y `gradient` de MATLAB para diferenciar la función $f(x) = 0.2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$ de $x = 0$ a 0.8. Compare sus resultados con la solución exacta: $f'(x) = 25 - 400x^2 + 2\,025x^2 - 3\,600x^3 + 2\,000x^4$.

Solución Se puede primero expresar $f(x)$ como una función anónima

```
>> f=@(x) 0.2+25*x-200*x.^2+675*x.^3-900*x.^4+400*x.^5;
```

Luego se genera una serie de valores igualmente espaciados de las variables independientes y dependientes,

```
>> x=0:0.1:0.8;
>> y=f(x);
```

La función `diff` es para determinar las diferencias entre elementos adyacentes de cada vector. Por ejemplo,

```
>> format short g
>> diff(x)

 0.1000 0.1000 0.1000 0.1000 0.1000 0.1000 0.1000 0.1000
```

Como se esperaba, el resultado representa la diferencia entre cada par de elementos de `x`. Para calcular aproximaciones de diferencia dividida de la derivada, simplemente se realiza una división vectorial de las diferencias de `y` entre las diferencias de `x` introduciendo

```
>> d=diff(y)./diff(x)

10.89 -0.01 3.19 8.49 8.69 1.39 -11.01 -21.31
```

Observe que, debido a que se están usando valores igualmente espaciados, después de generar los valores `x`, se podría simplemente haber realizado el cálculo anterior concisamente como

```
>> d=diff(f(x))/0.1;
```

El vector `d` contiene ahora estimados de derivada que corresponden al punto medio entre elementos adyacentes. Por lo tanto, con objeto de desarrollar una gráfica de los resultados, se debe primero generar un vector que contenga los valores `x` para el punto medio de cada intervalo.

```
>> n=length(x);
>> xm=(x(1:n-1)+x(2:n))./2;
```

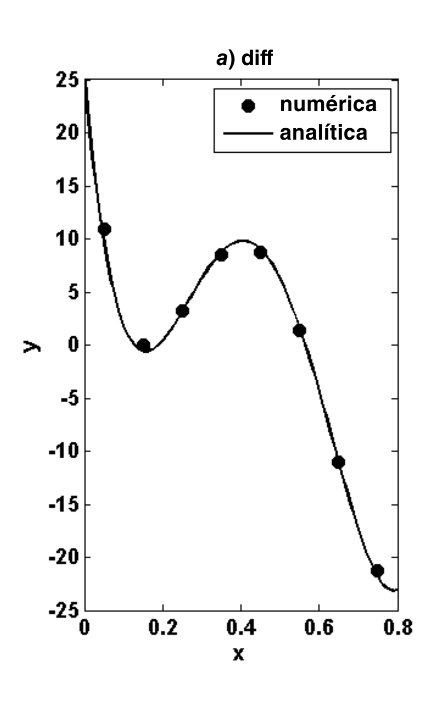
Se pueden calcular valores para la derivada analítica con un nivel más fino de resolución para incluirse en la gráfica para comparación

```
>> xa=0:.01:.8;
>> ya=25-400*xa+3*675*xa.^2-4*900*xa.^3+5*400*xa.^4;
```

Luego se genera una gráfica de los estimados numéricos y analíticos con

```
subplot(1,2,1), plot(xm,d,'o',xa,ya)
xlabel('x'),ylabel('y')
legend('numerical','analytical'),title('(a) diff')
```

Como se muestra en la figura 23.7*a*, los resultados de la aproximación numérica se comparan favorablemente con la solución analítica exacta para este caso.



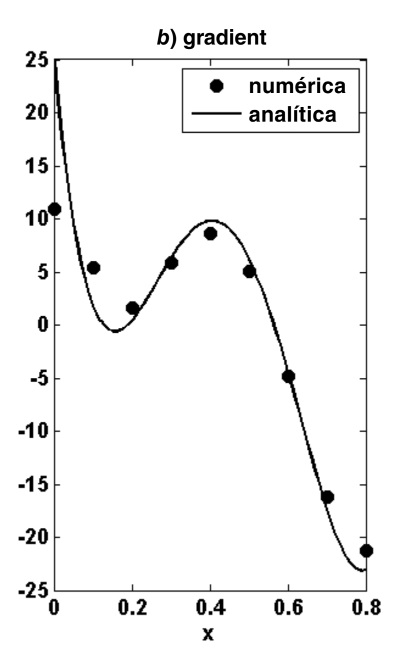


FIGURA 23.7
Comparación de la derivada exacta (línea) con estimados numéricos (círculos) calculada con las funciones *a*) `diff` y *b*) `gradient` de MATLAB.

También se puede usar la función `gradient` para determinar las derivadas como

```
>> dy=gradient(y,0.1)

dy = 10.89 5.44 1.59 5.84 8.59 5.04  -4.81 -16.16 -21.31
```

Como se hizo para la función `diff`, se puede también mostrar en una gráfica los estimados tanto numéricos como analíticos

```
>> subplot(1,2,2), plot(x,dy,'o',xa,ya)
>> xlabel('x')
>> legend('numerical','analytical'),title('(b) gradient')
```

Los resultados (figura 23.7*b*) no son tan exactos como los que se obtienen con la función `diff` (figura 23.7*a*). Esto se debe al hecho de que `gradient` utiliza intervalos que son el doble (0.2) de amplios que los que se usan para `diff` (0.1).

Más allá de los vectores unidimensionales, la función `gradient` es especialmente adecuada para determinar las derivadas parciales de matrices. Por ejemplo, para una matriz bidimensional, se puede llamar la función como

```
[fx, fy] = gradient(f, h)
```

donde *`f`* es una matriz bidimensional, *`fx`* corresponde a las diferencias en la dirección *x* (columnas) y *`fy`* corresponde a las diferencias en la dirección *y* (hileras), y *`h`* = espaciamiento entre puntos. Si se omite *`h`*, el espaciamiento entre puntos en ambas dimensiones se supone que es uno. En la sección 31.4.2 se ilustrará cómo se puede usar esta capacidad para visualizar campos vectoriales.

23.6.2 Mathcad

Mathcad tiene operadores que realizan integración y diferenciación numéricas. Estos operadores utilizan los mismos símbolos matemáticos tradicionales que usted ha empleado desde la preparatoria o su primer semestre en la universidad, y se ven como dichos símbolos.

El operador de integración usa una secuencia de evaluaciones de la integral por regla trapezoidal, y el algoritmo de Romberg. Se realizan iteraciones hasta que los resultados sucesivos varíen en

menos que la tolerancia. El operador de derivada utiliza un método similar para calcular derivadas de orden entre 0 y 5. Este operador crea una tabla de aproximaciones con base en cálculos de la derivada por diferencia dividida, usando varias órdenes y tamaños de paso. Se usan técnicas de extrapolación para estimar valores de un modo parecido al método de Richardson.

La figura 23.8 muestra un ejemplo de Mathcad donde f(x) se crea usando el símbolo de definición (:=) y luego se calcula la integral en un rango de x = 0 y x = 0.8. En este caso se usó el polinomio simple que se evaluó en el capítulo 21. Observe que la gama definida por las variables a y b se introduce con el símbolo de definición.

La figura 23.9 muestra un ejemplo de Mathcad donde una función f(x) se crea con el símbolo de definición (:=) y luego se calculan la primera y la tercera derivadas en un punto en el que x = –6. Observe que la ubicación del punto y el orden de la derivada se introducen con el símbolo de definición.

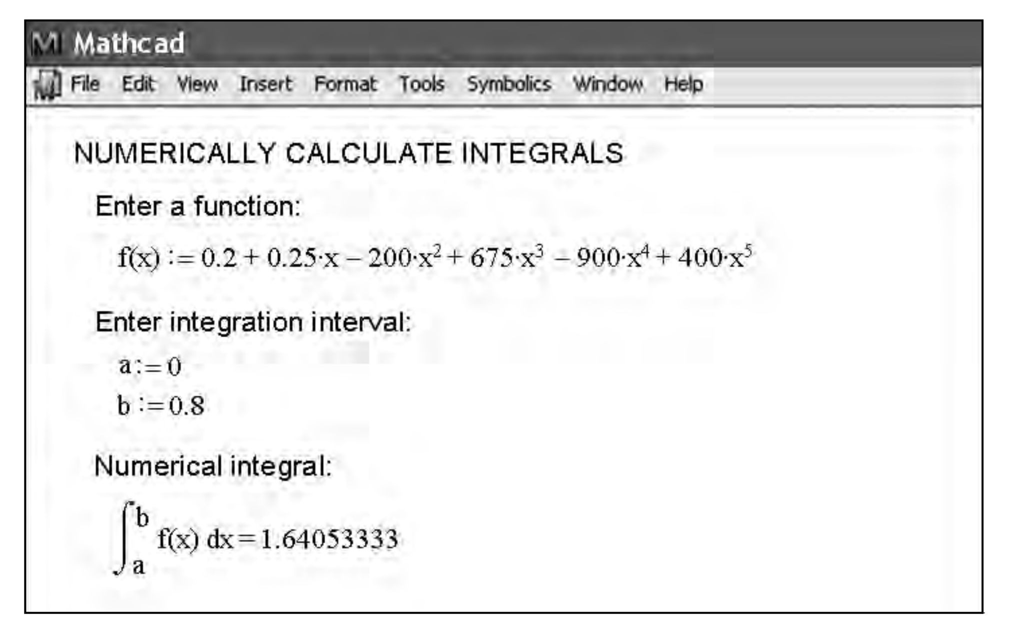


Figura 23.8
Pantalla Mathcad para determinar la integral de un polinomio con integración Romberg.

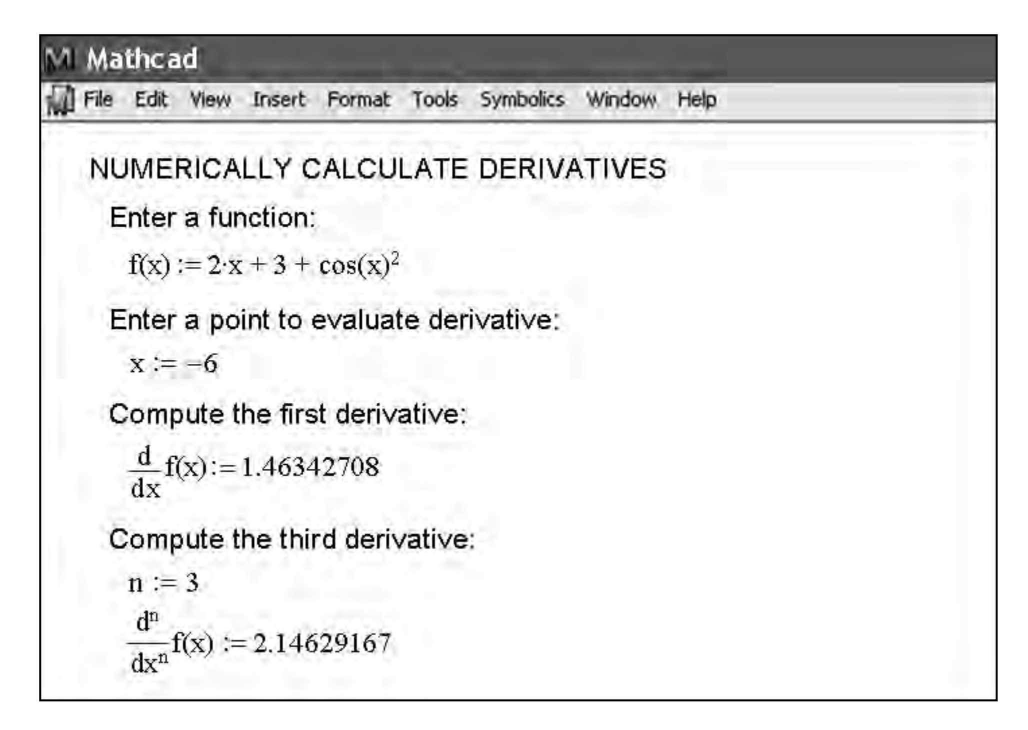


Figura 23.9
Pantalla Mathcad para implementar diferenciación numérica.

PROBLEMAS

23.1 Calcule las aproximaciones por diferencias hacia adelante y hacia atrás, de $O(h)$ y $O(h^2)$, y aproximaciones por diferencia central de $O(h^2)$ y $O(h^4)$ para la primera derivada de $y = \text{sen } x$, en $x = \pi/4$, con el uso de un valor de $h = \pi/12$. Estime el error relativo porcentual verdadero ε_t para cada aproximación.

23.2 Repita el problema 23.1, pero para $y = \log x$ evaluada en $x = 25$ con $h = 2$.

23.3 Use aproximaciones por diferencias centradas para estimar las derivadas primera y segunda de $y = e^x$ en $x = 2$ para $h = 0.1$. Emplee las dos fórmulas de $O(h^2)$ y $O(h^4)$ para hacer sus estimaciones.

23.4 Emplee la extrapolación de Richardson para estimar la primera derivada de $y = \cos x$ en $x = \pi/4$, con el uso de tamaños de paso de $h_1 = \pi/3$ y $h_2 = \pi/6$. Utilice diferencias centradas de $O(h^2)$ para las estimaciones iniciales.

23.5 Repita el problema 23.4, pero para la primera derivada de ln x en $x = 5$, con $h_1 = 2$ y $h_2 = 1$.

23.6 Emplee la ecuación (23.9) para determinar la primera derivada de $y = 2x^4 - 6x^3 - 12x - 8$ en $x = 0$, con base en los valores de $x_0 = -0.5$, $x_1 = 1$ y $x_2 = 2$. Compare este resultado con el valor verdadero y con una estimación obtenida con el uso de una aproximación por diferencias centradas con base en $h = 1$.

23.7 Demuestre que para puntos de datos equidistantes, la ecuación (23.9) se reduce a la ecuación (4.22) en $x = x_i$.

23.8 Calcule las aproximaciones por diferencia central de primer orden de $O(h^4)$ para cada una de las funciones siguientes en la ubicación y con el tamaño de paso que se especifica:

a) $y = x^3 + 4x - 15$ en $x = 0$, $h = 0.25$
b) $y = x^2 \cos x$ en $x = 0.4$, $h = 0.1$
c) $y = \tan (x/3)$ en $x = 3$, $h = 0.5$
d) $y = \text{sen } (0.5\sqrt{x})/x$ en $x = 1$, $h = 0.2$
e) $y = e^x + x$ en $x = 2$, $h = 0.2$

Compare sus resultados con las soluciones analíticas.

23.9 Para un cohete, se recabaron los datos siguientes de la distancia recorrida *versus* el tiempo:

t, s	0	25	50	75	100	125
y, km	0	32	58	78	92	100

Use diferenciación numérica para estimar la velocidad y aceleración del cohete en cada momento.

23.10 Desarrolle un programa amigable para el usuario a fin de aplicar el algoritmo de Romberg para estimar la derivada de una función dada.

23.11 Desarrolle un programa amigable para el usuario, que obtenga estimaciones de la primera derivada para datos irregularmente espaciados. Pruébelo con los datos siguientes:

x	1	1.5	1.6	2.5	3.5
f(*x*)	0.6767	0.3734	0.3261	0.08422	0.01596

donde $f(x) = 5e^{-2x}x$. Compare sus resultados con las derivadas verdaderas.

23.12 Se dan los siguientes datos para la velocidad de un objeto como función del tiempo:

t, s	0	4	8	12	16	20	24	28	32	36
v, m/s	0	34.7	61.8	82.8	99.2	112.0	121.9	129.7	135.7	140.4

a) Usando el mejor método numérico disponible, ¿qué tan lejos viaja el objeto desde $t = 0$ hasta 28 s?
b) Usando el mejor método numérico disponible, ¿cuál es la aceleración del objeto a $t = 28$ s?
c) Usando el mejor método numérico disponible, ¿cuál es la aceleración del objeto a $t = 0$ s?

23.13 Recuerde que para el problema del paracaidista que desciende, la velocidad está dada por

$$v(t) = \frac{gm}{c}\left(1 - e^{-(c/m)t}\right) \tag{P23.13.1}$$

y la distancia recorrida se obtiene con

$$d(t) = \frac{gm}{c}\int_0^t \left(1 - e^{-(c/m)t}\right) dt \tag{P23.13.2}$$

Dadas $g = 9.81$, $m = 70$ y $c = 12$,

a) Use MATLAB o Mathcad para integrar la ecuación (P23.13.1), de $t = 0$ a 10.
b) Integre la ecuación (P23.13.2) en forma analítica, con la condición inicial de que $d = 0$ y $t = 0$. Evalúe el resultado en $t = 10$ para confirmar el inciso *a*).
c) Emplee MATLAB o Mathcad para diferenciar la ecuación (P23.13.1) en $t = 10$.
d) Diferencie en forma analítica la ecuación (P23.13.1) en $t = 10$ para confirmar el inciso *c*).

23.14 La distribución normal se define como

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

a) Utilice MATLAB o Mathcad para integrar esta función de $x = -1$ a 1, y de -2 a 2.
b) Use MATLAB o Mathcad para determinar los puntos de inflexión de esta función.

23.15 Los datos siguientes se generaron a partir de la distribución normal:

x	–2	–1.5	–1	–0.5	0
f(*x*)	0.05399	0.12952	0.24197	0.35207	0.39894

x	0.5	1	1.5	2
f(*x*)	0.35207	0.24197	0.12952	0.05399

a) Utilice MATLAB para integrar estos datos de $x = -1$ a 1 y -2 a 2, con la función `trap`.
b) Emplee MATLAB para estimar los puntos de inflexión de estos datos.

23.16 Evalúe $\partial f/\partial x$, $\partial f/\partial y$ y $\partial f/(\partial x \partial y)$ para la siguiente función en $x = y = 1$ *a*) analíticamente y *b*) numéricamente $\Delta x = \Delta y = 0.0001$.

$$f(x, y) = 3xy + 3x - x^3 - 3y^3$$

23.17 Escriba un programa en MATLAB que integre

$$\int_0^{2/\pi} \frac{\text{sen } t}{t} dt$$

con el uso de las funciones `quad` y `quadl`. Para aprender más acerca de `quadl`, escriba `help quadl` en la barra de MATLAB.

23.18 Use el comando `diff` en MATLAB y calcule la aproximación por diferencia finita de la primera y segunda derivadas en cada valor de x de los que se muestran en la siguiente tabla, excepto los dos puntos extremos. Use aproximaciones por diferencias finitas que sean correctas en el segundo orden, $O(\Delta x^2)$.

x	0	1	2	3	4	5	6	7	8	9	10
y	1.4	2.1	3.3	4.8	6.8	6.6	8.6	7.5	8.9	10.9	10

23.19 El objetivo de este problema es comparar las aproximaciones por diferencias finitas de segundo orden exactas hacia delante, atrás y centradas, de la primera derivada de una función con el valor real de la derivada. Esto se hará para

$$f(x) = e^{-2x} - x$$

a) Use el cálculo para determinar el valor correcto de la derivada en $x = 2$.

b) Para evaluar las aproximaciones por diferencias finitas centradas, comience con $x = 0.5$. Así, para la primera evaluación, los valores de x para la aproximación por diferencias centradas será $x = 2 \pm 0.5$ o $x = 1.5$ y 2.5. Entonces, disminuya en pasos de 0.01 hacia abajo hasta un valor mínimo de $\Delta x = 0.01$.

c) Repita el inciso *b*) para las diferencias de segundo orden hacia delante y hacia atrás. (Observe que esto se puede hacer al mismo tiempo que la diferencia centrada se calcula en el ciclo o *loop*.)

d) Grafique los resultados de *b*) y *c*) *versus* x. Para efectos de comparación, incluya el resultado exacto en la gráfica.

23.20 Use una expansión en series de Taylor para obtener una aproximación a la tercera derivada que tenga una exactitud de segundo orden, por diferencias finitas centradas. Para hacer esto, tendrá que usar cuatro expansiones diferentes para los puntos x_{i-2}, x_{i-1}, x_{i+1} y x_{i+2}. En cada caso, la expansión será alrededor del punto x_i. El intervalo Δx se usará en cada caso de $i - 1$ e $i + 1$, y $2\Delta x$ se empleará en cada caso de $i - 2$ e $i + 2$. Las cuatro ecuaciones deben combinarse en forma que se elimine las derivadas primera y segunda. Utilice suficientes términos en cada expansión para evaluar el primer término que se truncará a fin de determinar el orden de la aproximación.

23.21 Use los datos siguientes para encontrar la velocidad y aceleración en $t = 10$ segundos:

Tiempo, t, s	0	2	4	6	8	10	12	14	16
Posición, x, m	0	0.7	1.8	3.4	5.1	6.3	7.3	8.0	8.4

Emplee los métodos de diferencias finitas correctas de segundo orden *a*) centradas, *b*) hacia adelante, *c*) hacia atrás.

23.22 Un avión es seguido por radar, y se toman datos cada segundo en coordenadas polares θ y r.

t, s	200	202	204	206	208	210
θ, (rad)	0.75	0.72	0.70	0.68	0.67	0.66
r, m	5 120	5 370	5 560	5 800	6 030	6 240

A los 206 segundos, utilice diferencias finitas centradas (correctas de segundo orden) para encontrar las expresiones vectoriales para la velocidad $\vec{v}$, y aceleración $\vec{a}$. La velocidad y aceleración en coordenadas polares son:

$$\vec{v} = \dot{r}\vec{e}_r + r\dot{\theta}\vec{e}_\theta \qquad \text{y} \qquad \vec{a} = (\ddot{r} - r\dot{\theta}^2)\vec{e}_r + (r\ddot{\theta} + 2\dot{r}\dot{\theta})\vec{e}_\theta$$

23.23 Desarrolle un programa de macros en Excel VBA para leer en columnas adyacentes de una hoja de cálculo los valores de x y y. Evalúe las derivadas en cada punto con el uso de la ecuación (23.9), y muestre los resultados en una tercera columna que se construya en la hoja, adyacente a las de los valores x y y. Pruebe su programa aplicándolo para evaluar las velocidades para los valores tiempo-posición del problema 23.21.

23.24 Use regresión para estimar la aceleración en cada momento para los datos siguientes con polinomios de segundo, tercero y cuarto grados. Grafique los resultados.

t	1	2	3.25	4.5	6	7	8	8.5	9.3	10
v	10	12	11	14	17	16	12	14	14	10

23.25 Usted tiene que medir la tasa de flujo de agua a través de un tubo pequeño. Para ello, coloque una boquilla en la salida del tubo y mida el volumen a través de ella como función del tiempo, según se ha tabulado a continuación. Estime la tasa de flujo en $t = 7$ s.

Tiempo, s	0	1	5	8
Volumen, cm^3	0	1	8	16.4

23.26 Se mide la velocidad v (m/s) del aire que fluye por una superficie plana a distintas distancias, y (m) de la superficie. Determine el esfuerzo cortante τ (N/m^2) en la superficie ($y = 0$) usando la *ley de viscosidad de Newton*.

$$\tau = \mu \frac{dv}{dy}$$

Suponga un valor de viscosidad dinámica $\mu = 1.8 \times 10^{-5}$ N · s/m^2.

y, m	0	0.002	0.006	0.0012	0.018	0.024
v, m/s	0	0.287	0.899	1.915	3.048	4.299

23.27 Es frecuente que las reacciones químicas sigan este modelo:

$$\frac{dc}{dt} = -kc^n$$

donde c = concentración, t = tiempo, k = tasa de reacción y n = orden de reacción. Es posible evaluar valores dados de c y dc/dt, k y n, por medio de regresión lineal del logaritmo de esta ecuación:

$$\log\left(-\frac{dc}{dt}\right) = \log k + n \log c$$

Use este enfoque y los datos que siguen para estimar los valores de k y n:

t	10	20	30	40	50	60
c	3.52	2.48	1.75	1.23	0.87	0.61

23.28 El perfil de velocidad de un fluido en un tubo circular se puede representar como

$$v = 10\left(1 - \frac{r}{r_0}\right)^{1/n}$$

donde v = velocidad, r = distancia radial medida desde la línea axial del tubo, r_0 = radio del tubo y n = parámetro. Determine el flujo en el tubo si $r_0 = 0.75$ y $n = 7$, usando *a*) integración de Romberg a una tolerancia de 0.1%, *b*) fórmula de dos puntos de Gauss-Legendre y *c*) la función `quad` de MATLAB. Observe que flujo = velocidad por área.

23.29 La cantidad de masa transportada por un tubo en un periodo dado se puede calcular como

$$M = \int_{t_1}^{t_2} Q(t)c(t)dt$$

donde M = masa (mg), t_1 = tiempo inicial (min), t_2 = tiempo final (min), $Q(t)$ = caudal (m^3/min) y $c(t)$ = concentración (mg/m^3). Las siguientes representaciones funcionales definen las variaciones temporales en flujo y concentración:

$$Q(t) = 9 + 4\cos^2(0.4t)$$
$$c(t) = 5e^{-0.5t} + 2e^{0.15t}$$

Determine la masa transportada entre $t_1 = 2$ y $t_2 = 8$ min, con *a*) integración de Romberg a una tolerancia de 0.1% y *b*) función `quad` de MATLAB.

CAPÍTULO 24

Estudio de casos: integración y diferenciación numéricas

El propósito del presente capítulo es aplicar los métodos de integración y diferenciación numérica, expuestos en la parte seis, a problemas prácticos de la ingeniería. Son dos situaciones que se presentan con mayor frecuencia. En el primer caso, la función que modela un problema puede tener una forma analítica demasiado complicada para resolverse con los métodos del cálculo. En situaciones de este tipo, se aplican los métodos numéricos usando la expresión analítica para generar una tabla de valores de la función. En el segundo caso, la función que habrá de usarse en el cálculo se halla en forma tabular. Este tipo de función generalmente representa una serie de mediciones, observaciones o alguna otra información empírica. Los datos para cualquiera de los casos son directamente compatibles con los esquemas analizados en esta parte del libro.

En la sección 24.1, que trata del cálculo de la cantidad de calor en la ingeniería química, se emplean ecuaciones. Aquí se integra numéricamente una función analítica para determinar el calor requerido para elevar la temperatura de un material.

Las secciones 24.2 y 24.3 también usan funciones que están en forma de ecuación. La sección 24.2, tomada de la ingeniería civil, emplea la integración numérica para determinar la fuerza del viento que actúa sobre el mástil de un velero de carreras. La sección 24.3 determina la raíz media cuadrática de la corriente para un circuito eléctrico. Este ejemplo sirve para demostrar la utilidad de la integración de Romberg y de la cuadratura de Gauss.

La sección 24.4 se concentra en el análisis de la información tabular para determinar el trabajo necesario para mover un bloque. Aunque esta aplicación tiene una vinculación directa con la ingeniería mecánica, se relaciona con todas las otras áreas de la ingeniería. Además, usamos este ejemplo para ilustrar la integración de datos irregularmente espaciados.

24.1 INTEGRACIÓN PARA DETERMINAR LA CANTIDAD TOTAL DE CALOR (INGENIERÍA QUÍMICA/BIOINGENIERÍA)

Antecedentes En ingeniería química y en bioingeniería se emplean cálculos de la cantidad de calor en forma rutinaria, así como en muchos otros campos de la ingeniería. Esta aplicación ofrece un ejemplo simple, pero útil, de tales cálculos.

La determinación de la cantidad de calor requerido para elevar la temperatura de un material es un problema con el que a menudo nos enfrentamos. La característica necesaria para llevar a cabo este cálculo es la capacidad calorífica c. Este parámetro representa la cantidad de calor requerida para elevar una unidad de masa en una unidad de temperatura. Si c es constante en el intervalo de temperaturas que se examinan, el calor requerido ΔH (en calorías) se calcula mediante

$$\Delta H = mc\,\Delta T \tag{24.1}$$

donde c tiene unidades de cal/(g · °C), m = masa (g) y ΔT = cambio de temperatura (°C). Por ejemplo, la cantidad de calor necesaria para elevar la temperatura de 20 gramos de agua desde 5 hasta 10°C es igual a:

$$\Delta H = 20(1)(10 - 5) = 100 \text{ cal}$$

donde la capacidad calorífica del agua es aproximadamente 1 cal/(g · °C). Este cálculo es adecuado cuando ΔT es pequeño. Sin embargo, para grandes cambios de temperatura, la capacidad calorífica no es constante y, de hecho, varía en función de la temperatura. Por ejemplo, la capacidad calorífica de un material podría aumentar con la temperatura de acuerdo con una relación tal como:

$$c(T) = 0.132 + 1.56 \times 10^{-4}T + 2.64 \times 10^{-7}T^2 \qquad (24.2)$$

En este caso se pide por ejemplo calcular el calor necesario para elevar la temperatura de 1000 gramos de este material desde –100 hasta 200°C.

Solución La ecuación (PT6.4) ofrece una manera para calcular el valor promedio de $c(T)$:

$$\bar{c}(T) = \frac{\int_{T_1}^{T_2} c(T)\, dT}{T_2 - T_1} \qquad (24.3)$$

que se sustituye en la ecuación (24.1) para dar:

$$\Delta H = m \int_{T_1}^{T_2} c(T)\, dT \qquad (24.4)$$

donde $\Delta T = T_2 - T_1$. Ahora como, en el caso actual, $c(T)$ es una función cuadrática, ΔH puede determinarse de manera analítica. La ecuación (24.2) se sustituye en la ecuación (24.4) y después se integra para dar un valor exacto, ΔH = 42 732 cal. Es útil e instructivo comparar este resultado con los métodos numéricos desarrollados en el capítulo 21. Para esto, es necesario generar una tabla de valores de c para distintos valores de T, que se muestra a la derecha.

T, °C	*c*, cal/(g · °C)
–100	0.11904
–50	0.12486
0	0.13200
50	0.14046
100	0.15024
150	0.16134
200	0.17376

Estos puntos se utilizan junto con una regla de Simpson 1/3 con seis segmentos calculándose una estimación de la integral de 42 732, este resultado se sustituye en la ecuación (24.4) para obtener un valor de ΔH = 42 732 cal, el cual concuerda exactamente con la solución analítica. Esta concordancia exacta ocurriría sin importar cuántos segmentos se utilicen. Se espera tal resultado debido a que c es una función cuadrática y la regla de Simpson es exacta para polinomios de tercer grado o menores (véase la sección 21.2.1).

Los resultados que se obtuvieron con la regla del trapecio se muestran en la tabla 24.1. Parece que la regla del trapecio es también capaz de estimar el calor total de manera exacta. No obstante, se requiere un paso pequeño (< 10°C) para una exactitud de cinco cifras. El mismo cálculo también se puede implementar con software. Por ejemplo, MATLAB da

Tabla 24.1 Resultados usando la regla del trapecio con varios tamaños de paso.

Tamaño de paso, °C	ΔH	ε_t(%)
300	96 048	125
150	43 029	0.7
100	42 864	0.3
50	42 765	0.07
25	42 740	0.018
10	42 733.3	<0.01
5	42 732.3	<0.01
1	42 732.01	<0.01
0.05	42 732.00003	<0.01

```
>> m=1000;
>> DH=m*quad(@(T) 0.132+1.56e-4*T+2.64e-7*T.^2,-100,200)
DH =
  42732
```

24.2 FUERZA EFECTIVA SOBRE EL MÁSTIL DE UN BOTE DE VELA DE CARRERAS (INGENIERÍA CIVIL/AMBIENTAL)

Antecedentes En la figura 24.1*a* se muestra la sección transversal de un bote de vela de carreras. Las fuerzas del viento (f), ejercidas por pie de mástil de las velas, varían en función de la distancia sobre la cubierta del bote (z), como se indica en la figura 24.1*b*. Calcule la fuerza de tensión T en el cable de soporte izquierdo del mástil, suponiendo que el cable de soporte derecho está totalmente flojo y que el mástil se une a la cubierta de modo que transmite fuerzas horizontales o verticales, pero no momentos. Suponga que el mástil permanece vertical.

Solución Para resolver este problema, se requiere que la fuerza distribuida f se convierta en una fuerza total equivalente F y que se calcule su localización d sobre la cubierta (figura 24.2). Este

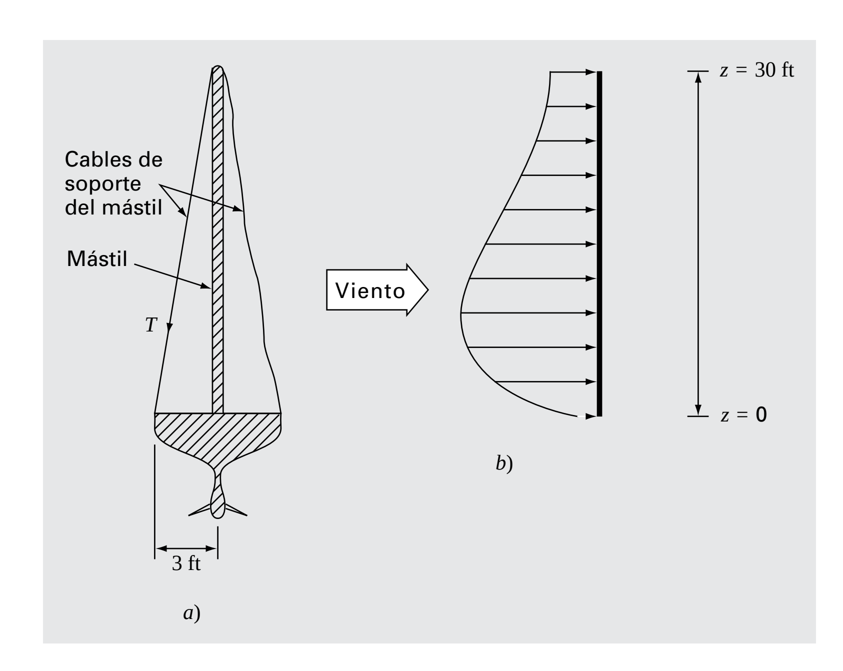


Figura 24.1
a) Sección transversal de un bote de vela de carreras. *b*) Fuerzas del viento *f* ejercidas por pie de mástil en función de la distancia *z* sobre la cubierta del bote.

cálculo se complica por el hecho de que la fuerza ejercida por pie de mástil varía con la distancia sobre la cubierta. La fuerza total ejercida sobre el mástil se expresa como la integral de una función continua:

$$F = \int_0^{30} 200\left(\frac{z}{5+z}\right) e^{-2z/30}\, dz \tag{24.5}$$

Esta integral no lineal es difícil de evaluar en forma analítica. Por lo tanto, en este problema es conveniente emplear procedimientos numéricos, como las reglas de Simpson y la del trapecio. Esto se lleva a cabo al calcular $f(z)$ para diferentes valores de z y después utilizar la ecuación (21.10) o la (21.18). Por ejemplo, la tabla 24.2 tiene valores de $f(z)$ para un tamaño de paso de 3 ft, que proporciona datos para la regla de Simpson 1/3 o para la regla del trapecio. Los resultados para varios tamaños de paso se muestran en la tabla 24.3. Se observa que ambos métodos dan un valor, F = 1 480.6 lb conforme el tamaño de paso se va haciendo pequeño. Aquí, los tamaños de paso de 0.05 ft para la regla del trapecio y de 0.5 ft para la regla de Simpson proporcionan buenos resultados.

Tabla 24.2 Valores de $f(z)$ para un tamaño de paso de 3 ft que proporcionan los datos para las reglas del trapecio y de Simpson 1/3.

z, ft	0	3	6	9	12	15	18	21	24	27	30
f(*z*),lb/ft	0	61.40	73.13	70.56	63.43	55.18	47.14	39.83	33.42	27.89	23.20

La distancia efectiva d en la que actúa la fuerza (figura 24.2) se calcula evaluando la integral:

$$d = \frac{\int_0^{30} z f(z)\, dz}{\int_0^{30} f(z)\, dz} \tag{24.6}$$

o

$$d = \frac{\int_0^{30} 200z[z/(5+z)]e^{-2z/30} dz}{1\,480.6} \tag{24.7}$$

Esta integral puede evaluarse usando métodos similares a los anteriores. Por ejemplo, la regla de Simpson 1/3 con tamaño de paso 0.5 da d = 19 326.9/1 480.6 = 13.05 ft.

Conocidos F y d mediante los métodos numéricos, ahora se emplea un diagrama de cuerpo libre para desarrollar ecuaciones de balance de fuerza y de momento. El diagrama

$\theta = \tan^{-1}(3/30) = 0.0996687$
F = 1 480.6 lb
d = 13.05 ft
T
0
3 ft
H
V

Figura 24.2
Diagrama de cuerpo libre de las fuerzas ejercidas sobre el mástil de un velero.

de cuerpo libre se muestra en la figura 24.2. Sumando fuerzas en las direcciones vertical y horizontal, y tomando momentos respecto al punto 0 se obtiene:

$$\sum F_H = 0 = F - T \operatorname{sen} \theta - H \tag{24.8}$$

$$\sum F_V = 0 = V - T \cos \theta \tag{24.9}$$

$$\sum M_0 = 0 = 3V - Fd \tag{24.10}$$

donde T = la tensión en el cable y H y V = las reacciones desconocidas sobre el mástil transmitidas por la cubierta. Tanto la dirección como la magnitud de H y V son desconocidas. De la ecuación (24.10) se despeja directamente V, puesto que se conocen F y d:

$$V = \frac{Fd}{3} = \frac{(1\,480.6)(13.05)}{3} = 6\,440.6 \text{ lb}$$

Por lo tanto, a partir de la ecuación (24.9):

$$T = \frac{V}{\cos \theta} = \frac{6\,440.6}{0.995} = 6\,473 \text{ lb}$$

y de la ecuación (24.8):

$$H = F - T \operatorname{sen} \theta = 1\,480.6 - (6\,473)(0.0995) = 836.54 \text{ lb}$$

Ahora el conocer estas fuerzas nos permite continuar con otros aspectos del diseño estructural del bote, como los cables y el sistema de soporte del mástil en la cubierta. Este problema ilustra claramente dos usos de la integración numérica que pueden encontrarse en el diseño de estructuras en ingeniería. Se ve que ambas reglas, la del trapecio y la de Simpson 1/3, son fáciles de aplicar y constituyen herramientas prácticas en la solución de problemas. La regla de Simpson 1/3 es más exacta que la del trapecio para el mismo tamaño de paso, por lo que a menudo se prefiere aquélla.

24.3 RAÍZ MEDIA CUADRÁTICA DE LA CORRIENTE MEDIANTE INTEGRACIÓN NUMÉRICA (INGENIERÍA ELÉCTRICA)

Antecedentes El valor promedio de una corriente eléctrica oscilante en un periodo puede ser cero. Por ejemplo, suponga que la corriente se describe por una senoide simple: $i(t) = \operatorname{sen}(2\pi t/T)$, donde T es el periodo. El valor promedio de esta función se determina mediante la siguiente ecuación:

Tabla 24.3 Valores de *F* calculados con base en las diferentes versiones de las reglas del trapecio y de Simpson 1/3.

Técnica	Tamaño de paso, ft	Segmentos	*F*, lb
Regla del trapecio	15	2	1 001.7
	10	3	1 222.3
	6	5	1 372.3
	3	10	1 450.8
	1	30	1 477.1
	0.5	60	1 479.7
	0.25	120	1 480.3
	0.1	300	1 480.5
	0.05	600	1 480.6
Regla de Simpson 1/3	15	2	1 219.6
	5	6	1 462.9
	3	10	1 476.9
	1	30	1 480.5
	0.5	60	1 480.6

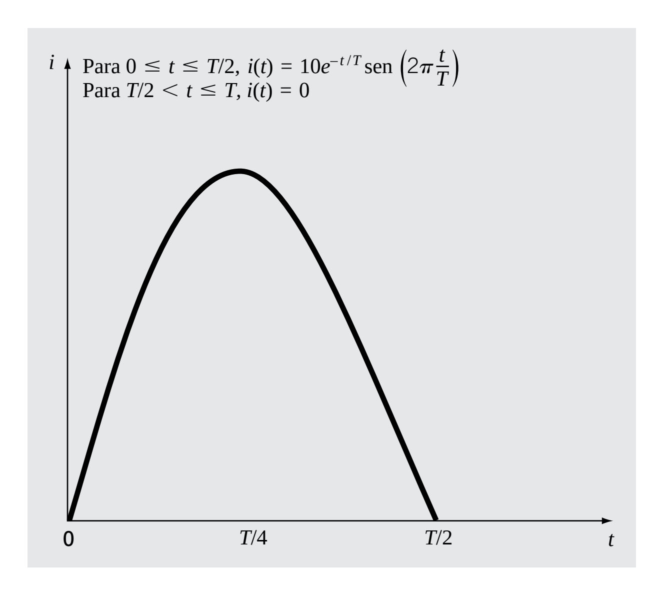


Figura 24.3
Una corriente eléctrica que varía en forma periódica.

$$i = \frac{\int_0^T \operatorname{sen}\left(\frac{2\pi t}{T}\right) dt}{T - 0} = \frac{-\cos\ (2\pi) + \cos 0}{T} = 0$$

A pesar del hecho de que el resultado total es cero, dicha corriente es capaz de realizar trabajo y generar calor. Por consiguiente, los ingenieros eléctricos a menudo caracterizan esa corriente por:

$$I_{\text{RMC}} = \sqrt{\frac{1}{T}\int_0^T i^2(t)\, dt} \tag{24.11}$$

donde $i(t)$ = la corriente instantánea. Calcule la RMC o raíz media cuadrática para la corriente que tiene la forma de onda mostrada en la figura 24.3, mediante la regla del trapecio, la regla de Simpson 1/3, la integración de Romberg y la cuadratura de Gauss para $T = 1$ s.

Solución En la tabla 24.4 se presentan las estimaciones de la integral para varias aplicaciones de la regla del trapecio y de la regla de Simpson 1/3. Observe que la regla de Simpson es más exacta que la del trapecio.

El valor exacto de la integral es 15.41261. Este resultado se obtiene utilizando la regla del trapecio con 128 segmentos, o la regla de Simpson con 32 segmentos. Se determina también la misma estimación usando la integración de Romberg (figura 24.4).

Además, la cuadratura de Gauss también se utiliza para hacer la misma estimación. La determinación de la *raíz media cuadrática* para la corriente implica la evaluación de la integral (con $T = 1$):

$$I = \int_0^{1/2} (10e^{-t}\operatorname{sen} 2\pi t)^2\ dt \tag{24.12}$$

Primero, se efectúa un cambio de variable aplicando las ecuaciones (22.29) y (22.30) para obtener:

$$t = \frac{1}{4} + \frac{1}{4}t_d \qquad dt = \frac{1}{4}dt_d$$

Esas relaciones se sustituyen en la ecuación (24.12) para obtener:

$$I = \int_{-1}^{1}\left[10e^{-[1/4+(1/4)t_d]}\ \operatorname{sen} 2\pi\left(\frac{1}{4} + \frac{1}{4}t_d\right)\right]^2 \frac{1}{4}dt \tag{24.13}$$

En la fórmula de Gauss-Legendre con dos puntos, esta función se evalúa en $t_d = -1/\sqrt{3}$ y $1/\sqrt{3}$, y los resultados son 7.684096 y 4.313728, respectivamente. Tales valores se sustituyen en la

***O*(*h*²)**	***O*(*h*⁴)**	***O*(*h*⁶)**	***O*(*h*⁸)**	***O*(*h*¹⁰)**	***O*(*h*¹²)**
0	20.21769	15.16503	15.41502	15.41261	15.41261
15.16327	15.48082	15.41111	15.41262	15.41261	
15.40143	15.41547	15.41225	15.41261		
15.41196	15.41277	15.41261			
15.41257	15.41262				
15.41261					

Figura 24.4
Resultados al usar la integración de Romberg para estimar la RMC de la corriente.

Tabla 24.4 Valores para la integral calculados con diversos esquemas numéricos. El error relativo porcentual ε_t se basa en el valor verdadero, 15.41261.

Técnica	Segmentos	Integral	ε_t(%)
Regla del trapecio	1	0.0	100
	2	15.16327	1.62
	4	15.40143	0.0725
	8	15.41196	4.21×10^{-3}
	16	15.41257	2.59×10^{-4}
	32	15.41261	1.62×10^{-5}
	64	15.41261	1.30×10^{-6}
	128	15.41261	0
Regla de Simpson 1/3	2	20.21769	–31.2
	4	15.48082	–0.443
	8	15.41547	–0.0186
	16	15.41277	1.06×10^{-3}
	32	15.41261	0

ecuación (22.23) para obtener un estimado de la integral de 11.99782, que representa un error de $\varepsilon_t = 22.1\%$.

La fórmula de tres puntos es (tabla 22.1):

$$I = 0.5555556(1.237449) + 0.8888889(15.16327) + 0.5555556(2.684915)$$
$$= 15.65755 \qquad |\varepsilon_t| = 1.6\%$$

Los resultados al emplear las fórmulas con más puntos se resumen en la tabla 24.5.

La estimación de la integral, 15.41261, se sustituye en la ecuación (24.12) para calcular $I_{RMC} = 3.925890$ A. El resultado sirve, como guía para otros aspectos del diseño y la operación del circuito.

Tabla 24.5 Resultados al usar las fórmulas de la cuadratura de Gauss con varios puntos para estimar la integral.

Puntos	Estimación	ε_t(%)
2	11.9978243	22.1
3	15.6575502	–1.59
4	15.4058023	4.42×10^{-2}
5	15.4126391	-2.01×10^{-4}
6	15.4126109	-1.82×10^{-5}

24.4 INTEGRACIÓN NUMÉRICA PARA CALCULAR EL TRABAJO (INGENIERÍA MECÁNICA/AERONÁUTICA)

Antecedentes En ingeniería muchos problemas implican el cálculo del trabajo. La fórmula general es:

Trabajo = fuerza × distancia

Cuando se le presentó este concepto en sus cursos de física en el nivel medio superior, se le mostraron algunas aplicaciones simples mediante el uso de fuerzas que permanecían constantes durante todo el desplazamiento. Por ejemplo, si una fuerza de 10 lb se usaba para jalar un bloque a través de una distancia de 15 ft, el trabajo que se obtiene con esta fórmula es 150 ft · lb.

Aunque ese simple cálculo es útil para presentar el concepto, la solución de problemas reales por lo común es más complicada. Por ejemplo, suponga que la fuerza varía durante el proceso del cálculo. En tales casos, la ecuación para el trabajo ahora se expresa como

$$W = \int_{x_0}^{x_n} F(x)\, dx \tag{24.14}$$

donde W = trabajo (ft · lb), x_0 y x_n = las posiciones inicial y final, respectivamente, y $F(x)$ es una fuerza que varía con la posición. Si $F(x)$ es fácil de integrar, la ecuación (24.14) se puede resolver en forma analítica. No obstante, en la solución de un problema real, quizá la fuerza no se exprese de esa manera. De hecho, cuando se analizan los datos obtenidos de mediciones, la fuerza podría estar disponible sólo en forma tabular. En tales casos, la integración numérica es la única opción viable para la evaluación.

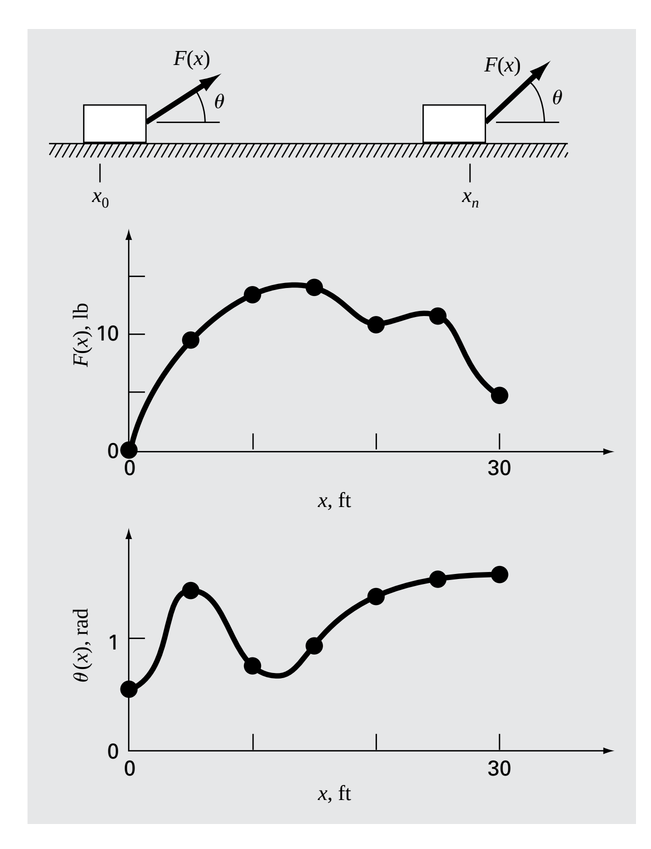


Figura 24.5
El caso de una fuerza variable que actúa sobre un bloque. Para este caso, tanto el ángulo como la magnitud de la fuerza varían.

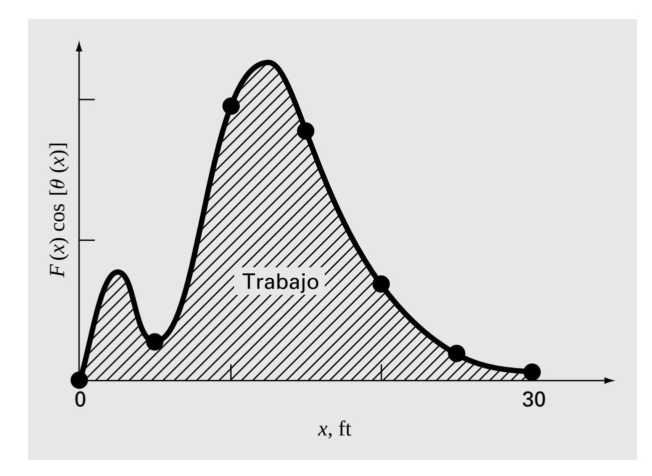


Figura 24.6
Una gráfica continua de $F(x)\cos[\theta(x)]$ contra la posición con los siete puntos discretos usados para las estimaciones de la integral numérica dadas en la tabla 24.7. Observe cómo el uso de los siete puntos para caracterizar esta función que varía en forma continua deja fuera dos picos en $x = 2.5$ y 12.5 ft.

Se obtiene mayor complejidad si el ángulo entre la fuerza y la dirección del movimiento también varía en función de la posición (figura 24.5). La ecuación del trabajo llega a dificultarse aún más al tomar en cuenta este efecto, entonces

$$W = \int_{x_0}^{x_n} F(x)\cos[\theta(x)]\,dx \tag{24.15}$$

De nuevo, si $F(x)$ y $\theta(x)$ son funciones sencillas, la ecuación (24.15) se podría resolver de manera analítica. Sin embargo, como se representa en la figura 24.5, es más común que la relación funcional sea complicada. En tal situación, los métodos numéricos ofrecen la única alternativa para determinar la integral.

Suponga que usted debe realizar el cálculo para la situación que se muestra en la figura 24.5. Aunque la figura indica los valores continuos de $F(x)$ y $\theta(x)$, considere que, debido a las restricciones experimentales, usted cuenta sólo con mediciones discretas a intervalos de $x = 5$ ft (tabla 24.6). Utilice versiones de una y múltiples aplicaciones de la regla del trapecio, y de las reglas de Simpson 1/3 y 3/8 para calcular el trabajo con estos datos.

Solución Los resultados del análisis se resumen en la tabla 24.7. Se calculó un error relativo porcentual ε_t con referencia al valor verdadero de la integral, 129.52, cuya estimación se realizó con base en los valores tomados de la figura 24.5 a intervalos de 1 ft.

Los resultados son interesantes, puesto que la mayor exactitud se obtiene en una aplicación de la regla del trapecio con dos segmentos. Las estimaciones más refinadas que utilizan más segmentos, así como las reglas de Simpson, dan resultados menos exactos.

La razón de este resultado, ilógico en apariencia, es porque el espaciamiento de los puntos no es el adecuado para captar las variaciones de las fuerzas y de los ángulos, lo cual es evidente en la figura 24.6, donde graficamos la curva continua del producto de $F(x)$ por $\cos[\theta(x)]$. Observe cómo el uso de siete puntos para caracterizar la variación continua de la función omite dos picos en $x = 2.5$ y 12.5 ft. La omisión de estos dos puntos efectivamente limita la exactitud de la estimación de la integración numérica dada en la tabla 24.7. El hecho de que la regla del trapecio con dos segmentos dé el resultado más exacto se debe a la posición casual de los puntos usados en este problema específico (figura 24.7).

Tabla 24.6 Datos para la fuerza $F(x)$ y el ángulo $\theta(x)$ como función de la posición x.

***x*, ft**	***F*(*x*), lb**	**θ, rad**	***F*(*x*) cos θ**
0	0.0	0.50	0.0000
5	9.0	1.40	1.5297
10	13.0	0.75	9.5120
15	14.0	0.90	8.7025
20	10.5	1.30	2.8087
25	12.0	1.48	1.0881
30	5.0	1.50	0.3537

Tabla 24.7 Estimaciones del trabajo calculado usando la regla del trapecio y las reglas de Simpson. El error relativo porcentual ε_t, como se calculó con referencia a un valor verdadero de la integral (129.52 lb · ft), se estimó con base en los valores en intervalos de 1 ft.

Técnica	Segmentos	Trabajo	ε_t,%
Regla del trapecio	1	5.31	95.9
	2	133.19	2.84
	3	124.98	3.51
	6	119.09	8.05
Regla de Simpson 1/3	2	175.82	–35.75
	6	117.13	9.57
Regla de Simpson 3/8	3	139.93	–8.04

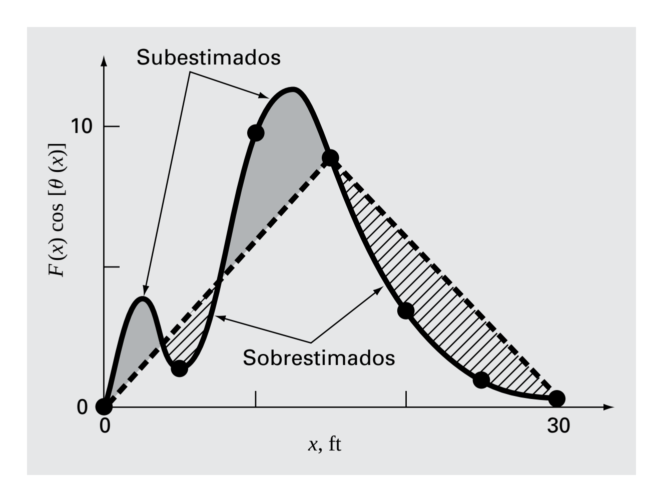


Figura 24.7
Representación gráfica del porqué la regla del trapecio con dos segmentos da una buena estimación de la integral en este caso específico. Casualmente, el uso de dos trapezoides lleva a un equilibrio entre los errores negativo y positivo.

La conclusión a partir de la figura 24.6 es que debe realizarse un número adecuado de mediciones para calcular las integrales con exactitud. En el presente caso si se tuvieran los datos en $F(2.5) \cos [\theta(2.5)] = 4.3500$ y $F(12.5) \cos [\theta(12.5)] = 11.3600$, podríamos determinar una estimación de la integral utilizando el algoritmo para datos irregularmente espaciados que se describió en la sección 21.3. La figura 24.8 ilustra la segmentación irregular en este caso. Si se incluyen dos puntos adicionales se obtiene una mejor estimación de la integral: 126.9 (ε_t = 2.02%). Así, la inclusión de datos adicionales incorporaría los picos que antes no se tomaron en cuenta y, en consecuencia, se tendrían mejores resultados.

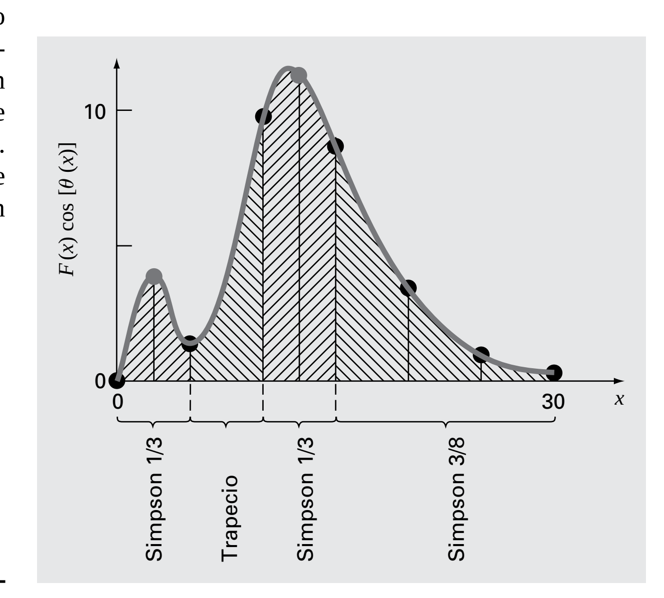


Figura 24.8
El esquema con segmentos irregulares que resulta de incluir dos puntos adicionales en x = 2.5 y 12.5 a los datos de la tabla 24.6. Se muestra la aplicación de las fórmulas de integración numérica a cada conjunto de segmentos.

PROBLEMAS

Ingeniería química/biológica

24.1 Realice el mismo cálculo que en la sección 24.1, pero obtenga la cantidad de calor requerido para elevar la temperatura de 1200 g del material, de –150 a 100°C. Use la regla de Simpson para hacer su cálculo, con valores de T en incrementos de 50°C.

24.2 Repita el problema 24.1, pero utilice la integración de Romberg con ε_s = 0.01%.

24.3 Vuelva a hacer el problema 24.1, pero emplee una fórmula de Gauss-Legendre de dos y tres puntos. Interprete sus resultados.

24.4 La integración proporciona un medio de calcular cuánta masa entra o sale de un reactor durante un periodo específico de tiempo, así

$$M = \int_{t_1}^{t_2} Qc\, dt$$

donde t_1 y t_2 = tiempos inicial y final, respectivamente. Esta fórmula es de sentido común si se recuerda la analogía entre la integración y la suma. Es decir, la integral representa la suma del producto del flujo por la concentración, lo que da la masa total que entra o sale de t_1 a t_2. Si la tasa de flujo es constante, Q se puede sacar de la integral:

$$M = Q\int_{t_1}^{t_2} c\, dt \qquad \text{(P24.4.1)}$$

Utilice integración numérica para evaluar esta ecuación para los datos que se enlistan a continuación. Observe que $Q = 4$ m^3/min.

t, min	0	10	20	30	35	40	45	50
c, mg/m^3	10	35	55	52	40	37	32	34

24.5 Use integración numérica para calcular cuánta masa sale de un reactor con base en las siguientes mediciones:

t, min	0	10	20	30	35	40	45	50
Q, m^3/min	4	4.8	5.2	5.0	4.6	4.3	4.3	5.0
c, mg/m^3	10	35	55	52	40	37	32	34

24.6 La *primera ley de la difusión de Fick* establece que

$$\text{Flujo de masa} = -D\frac{dc}{dx} \qquad \text{(P24.6.1)}$$

donde el flujo de masa = cantidad de masa que pasa a través de una unidad de área por unidad de tiempo (g/cm^2/s), D = coeficiente de difusión (cm^2/s), c = concentración y x = distancia (cm). Un ingeniero ambiental mide la concentración, que se presenta a continuación, de un contaminante en los sedimentos en el fondo de un lago ($x = 0$ en la interfase sedimento-agua y aumenta hacia abajo):

x, cm	0	1	3
c, 10^{-6} g/cm^3	0.06	0.32	0.6

Utilice la mejor técnica numérica de diferenciación disponible para estimar la derivada en $x = 0$. Emplee esta estimación junto con la ecuación (P24.6.1) para calcular el flujo de masa del contaminante que se desprende de los sedimentos hacia las aguas superiores ($D = 1.52 \times 10^{-6}$ cm^2/s). Para un lago con 3.6×10^6 m^2 de sedimentos, ¿cuánto contaminante será transportado hacia el lago durante un año?

24.7 Los siguientes datos se obtuvieron al cargar un gran buque petrolero:

t, min	0	10	20	30	45	60	75
V, 10^6 barriles	0.4	0.7	0.77	0.88	1.05	1.17	1.35

Calcule la tasa de flujo Q (es decir, dV/dt) para cada tiempo con un orden de h^2.

24.8 Usted está interesado en medir la velocidad de un fluido a través de un canal rectangular angosto abierto que conduce desperdicios de petróleo entre distintos lugares de una refinería. Usted sabe que, debido a la fricción con el fondo, la velocidad varía con la profundidad del canal. Si los técnicos sólo disponen de tiempo para hacer dos mediciones de la velocidad, ¿a qué profundidades las haría para obtener la mejor estimación de la velocidad promedio? Elabore recomendaciones en términos del porcentaje total de profundidad d medida a partir de la superficie del fluido. Por ejemplo, si se midiera en la superficie se tendría 0%d, mientras que en el fondo sería 100%d.

24.9 El tejido suave sigue una deformación de comportamiento exponencial ante la tensión uniaxial, mientras se encuentre en el rango fisiológico o normal de elongación. Esto se expresaría así:

$$\sigma = \frac{E_o}{a}(e^{a\varepsilon} - 1)$$

donde σ = esfuerzo, ε = tensión, y E_o y a son constantes materiales que se determinan en forma experimental. Para evaluar las dos constantes materiales, se deriva la ecuación anterior con respecto a ε, la cual es una relación fundamental para el tejido suave:

$$\frac{d\sigma}{d\varepsilon} = E_o + a\sigma$$

Para evaluar E_o y a, se emplean datos de esfuerzo-tensión para graficar $d\sigma/d\varepsilon$ *versus* σ, y la pendiente e intersección de esta gráfica son las dos constantes del material, respectivamente. La tabla siguiente contiene datos de esfuerzo-tensión para las cuerdas tendinosas del corazón (tendones pequeños que durante la contracción del músculo cardiaco mantienen cerradas sus válvulas). Éstos son datos tomados durante la carga del tejido, se obtendrían curvas distintas durante la descarga.

$\sigma \times 10^3$ N/m^2	87.8	96.6	176	263	350	569	833
$\varepsilon \times 10^{-3}$ m/m	153	198	270	320	355	410	460

$\sigma \times 10^3$ N/m^2	1 227	1 623	2 105	2 677	3 378	4 257
$\varepsilon \times 10^{-3}$ m/m	512	562	614	664	716	766

a) Calcule la derivada de $d\sigma/d\varepsilon$ por medio de diferencias finitas con exactitud de segundo orden. Grafique los datos y elimine aquellos puntos cerca de cero que parezcan no seguir la relación de línea recta. El error en dichos datos proviene de la incapacidad de los instrumentos para medir los valores pequeños en dicha región. Lleve a cabo un análisis de regresión de los demás puntos para determinar los valores de E_0 y a. Grafique los datos de esfuerzo *versus* tensión junto con la curva analítica expresada por la primera ecuación. Esto indicará qué tan bien se ajustan los datos a la curva analítica.

b) Es frecuente que el análisis anterior no funcione bien debido a que es difícil evaluar el valor de E_o. Para resolver este problema, no se utiliza E_o. Se selecciona un punto $(\bar{\sigma}, \bar{\varepsilon})$ de los datos que esté a la mitad del rango empleado para el análisis de regresión. Estos valores se sustituyen en la primera ecuación y se determina un valor E_0/a que se reemplaza en la primera ecuación:

$$\sigma = \left(\frac{\bar{\sigma}}{e^{a\bar{\varepsilon}} - 1}\right)(e^{a\bar{\varepsilon}} - 1)$$

Con este enfoque, los datos experimentales que están bien definidos producirán un buen ajuste entre los datos y la curva analítica. Emplee esta nueva relación y grafique otra vez los datos del esfuerzo *versus* la tensión y también la nueva curva analítica.

24.10 La técnica estándar para determinar el gasto cardiaco es el método de dilución de un colorante, desarrollado por Hamilton. Se inserta el extremo de un catéter pequeño en la arteria radial, y el otro se conecta a un densitómetro, que registra en forma automática la concentración del colorante en la sangre. Se inyectó con rapidez una cantidad conocida, 5.6 mg, de colorante y se obtuvieron los datos siguientes:

Tiempo, s	**Concentración, mg/L**	**Tiempo, s**	**Concentración, mg/L**
5	0	21	2.3
7	0.1	23	1.1
9	0.11	25	0.9
11	0.4	27	1.75
13	4.1	29	2.06
15	9.1	31	2.25
17	8	33	2.32
19	4.2	35	2.43

Al graficarse los datos anteriores se obtiene la curva de dilución del colorante que se muestra en la figura P24.10*a*. La concentración alcanza un valor máximo alrededor de 15 segundos después, luego hay

una disminución seguida de un aumento ocasionado por la recirculación del colorante. En la figura P24.10*b*, se muestra la curva graficada en escala semilogarítmica. Observe que la rama descendente de la curva de dilución se aproxima a una línea recta. A fin de separar el efecto de recirculación, los analistas extienden la porción de la línea recta. Entonces, el gasto cardiaco se calcula por medio de la ecuación siguiente:

$$C = \frac{M}{A} \times 60 \text{ s/min}$$

donde C = gasto cardiaco (L/min), M = cantidad de colorante inyectado (mg) y A = área bajo la curva con la corrección lineal. Calcule el gasto cardiaco de este paciente con el empleo de la regla del trapecio con un tamaño de paso de 2 s.

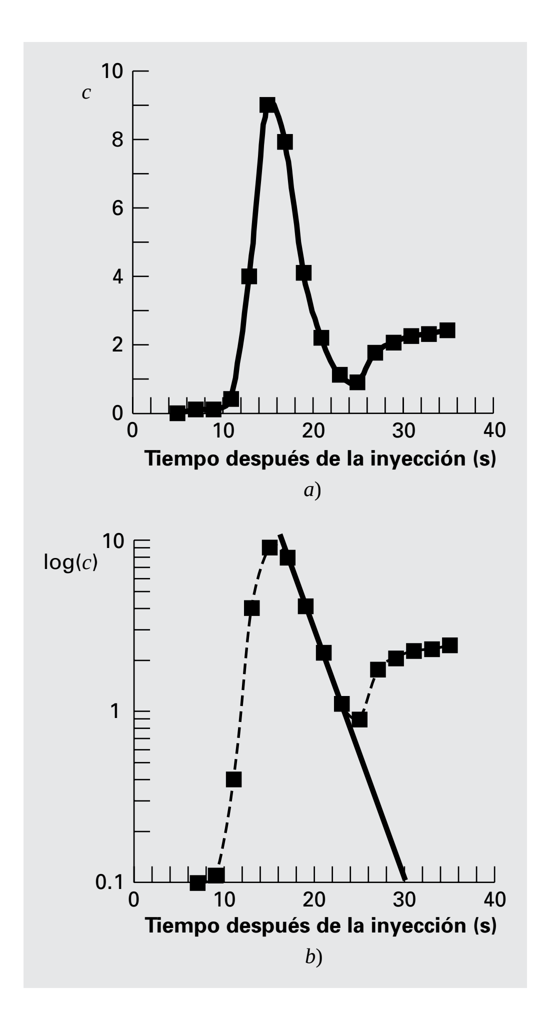


Figura P24.10

24.11 En todo el mundo, el glaucoma es la segunda causa principal de pérdida de la vista. La presión intraocular alta (presión dentro del ojo) casi siempre acompaña la pérdida de la visión. Existe la hipótesis de que la presión elevada daña un subconjunto de células en el ojo responsables de la vista. Un investigador postula que la relación entre la pérdida de la visión y la presión está descrita por la ecuación:

$$VL = A \exp\left(k \int_{25}^{t} (P - 13)\, dt\right)$$

donde *VL* es el porcentaje de pérdida de visión, *P* es la presión intraocular (mm de mercurio [mm Hg]), *t* es el tiempo (años), y *k* y *A* son constantes. Con el uso de los datos siguientes procedentes de tres pacientes, estime los valores de las constantes *k* y *A*.

Paciente	A	
Edad al emitir el diagnóstico	65	
VL	60	
	Edad, años	***P*, mm Hg**
	25	13
	40	15
	50	22
	60	23
	65	24

Paciente	B		C	
Edad al emitir el diagnóstico	43		80	
VL	40		30	
	Edad, años	***P*, mm Hg**	**Edad, años**	***P*, mm Hg**
	25	11	25	13
	40	30	40	14
	41	32	50	15
	42	33	60	17
	43	35	80	19

24.12 Una de sus colegas diseñó un parche transdérmico nuevo para aplicar insulina a través de la piel de los pacientes diabéticos en forma controlada, con lo que se elimina la necesidad de inyecciones dolorosas. Recabó los datos siguientes acerca del flujo de masa de la insulina que se aplica a través del parche (y piel) como función del tiempo:

Flujo, mg/cm²/h	Tiempo, h	Flujo, mg/cm²/h	Tiempo, h
15	0	8	5
14	1	5	10
12	2	2.5	15
11	3	2	20
9	4	1	24

Recuerde que el flujo de masa es la tasa de flujo a través de un área, o $(1/A)dm/dt$. Proporcione su mejor estimación posible de la cantidad de medicina distribuida a través de la piel en 24 horas de uso de un parche de 12 cm^2.

24.13 Se emplea la *videoangiografía* para medir el flujo sanguíneo y determinar el estado de la función circulatoria. A fin de cuantificar los videoangiogramas, se necesita conocer el diámetro del vaso sanguíneo y la velocidad de la sangre, de modo que se determine el flujo total de la sangre. A continuación se presenta el perfil densitométrico tomado de un videoangiograma de cierto vaso sanguíneo. Una forma de determinar de modo consistente a qué distancia del angiograma se localiza el borde del vaso sanguíneo, es determinar la primera derivada del perfil en un valor extremo. Con los datos que se proporcionan, encuentre las fronteras del vaso sanguíneo y estime el diámetro de éste. Emplee fórmulas de diferencias centradas tanto de $O(h^2)$ como de $O(h^4)$ y compare los resultados.

Distancia	Densidad	Distancia	Densidad
0	26.013	28	38.273
4	26.955	32	39.103
8	26.351	36	39.025
12	28.343	40	39.432
16	31.100	44	39.163
20	34.667	48	38.920
24	37.251	52	38.631

Distancia	Densidad	Distancia	Densidad
56	39.124	84	37.331
60	38.813	88	35.980
64	38.925	92	31.936
68	38.804	96	28.843
72	38.806	100	26.309
76	38.666	104	26.146
80	38.658		

Ingeniería civil /ambiental

24.14 Ejecute el mismo cálculo que en la sección 24.2, pero utilice integración de Romberg $O(h^8)$ para evaluar la integral.

24.15 Lleve a cabo el mismo cálculo que en la sección 24.2, pero emplee la cuadratura de Gauss para evaluar la integral.

24.16 Igual que en la sección 24.2, calcule el valor de F con el uso de la regla del trapecio y las de Simpson 1/3 y 3/8, pero utilice la fuerza siguiente. Divida el mástil en intervalos de cinco pies.

$$F = \int_0^{30} \frac{250z}{6+z} e^{-z/10}\, dz$$

24.17 Las áreas (A) de la sección transversal de una corriente se requieren para varias tareas de la ingeniería de recursos hidráulicos, como el pronóstico del escurrimiento y el diseño de presas. A menos que se disponga de dispositivos electrónicos muy avanzados para obtener perfiles continuos del fondo del canal, el ingeniero debe basarse en mediciones discretas de la profundidad para calcular A. En la figura P24.17 se representa un ejemplo de sección transversal común de una corriente. Los puntos de los datos representan ubicaciones en las que ancló un barco y se hicieron mediciones de la profundidad. Utilice aplicaciones (h = 4 y 2 m) de la regla del trapecio y de la de Simpson 1/3 (h = 2 m) para estimar el área de la sección transversal representada por esos datos.

24.18 Como se dijo en el problema 24.17, el área de la sección transversal de un canal se calcula con:

$$A_c = \int_0^B H(y)dy$$

donde B = ancho total del canal (m), H = profundidad (m) y y = distancia desde uno de los márgenes (m). En forma similar, el flujo promedio Q (m^3/s) se calcula por medio de:

$$Q = \int_0^B U(y)H(y)dy\,.$$

donde U = velocidad del agua (m/s). Use estas relaciones y algún método numérico para determinar A_c y Q, para los datos siguientes:

y, m	0	2	4	5	6	9
H, m	0.5	1.3	1.25	1.7	1	0.25
U, m/s	0.03	0.06	0.05	0.12	0.11	0.02

24.19 Se pueden usar las siguientes relaciones para analizar vigas uniformes sujetas a cargas distribuidas:

$$\frac{dy}{dx} = \theta(x) \quad \frac{d\theta}{dx} = \frac{M(x)}{EI} \quad \frac{dM}{dx} = V(x) \quad \frac{dV}{dx} = -w(x)$$

donde x = distancia a lo largo de la viga (m), y = deflexión (m), $\theta(x)$ = pendiente (m/m), E = módulo de elasticidad (Pa = N/m^2), I = momento de inercia (m^4), $M(x)$ = momento (N m), $V(x)$ = cizalladura (N) y $w(x)$ = carga distribuida (N/m). Para el caso de una carga linealmente creciente (recuerde la figura P8.18), la pendiente se puede calcular analíticamente como

$$\theta(x) = \frac{w_0}{120EIL}(-5x^4 + 6L^2x^2 - L^4) \qquad \text{(P24.19.1)}$$

Use *a*) integración numérica para calcular la deflexión (en m), y *b*) diferenciación numérica para calcular el momento (en N m) y la cizalladura (en N). Base sus cálculos numéricos en valores de la pendiente calculados con la ecuación (P24.19.1) a intervalos igualmente espaciados de Δx = 0.125 m a lo largo de una viga de 3 m. Use los siguientes valores de parámetros en su cálculo: E = 200 GPa, I = 0.0003 m^4 y w_0 = 2.5 kN/cm. Además, las deflexiones en los extremos de la viga se establecen en $y(0) = y(L) = 0$. Tenga cuidado con las unidades.

24.20 Usted mide las siguientes deflexiones a lo largo de una viga uniforme con un solo apoyo (véase el problema 24.19).

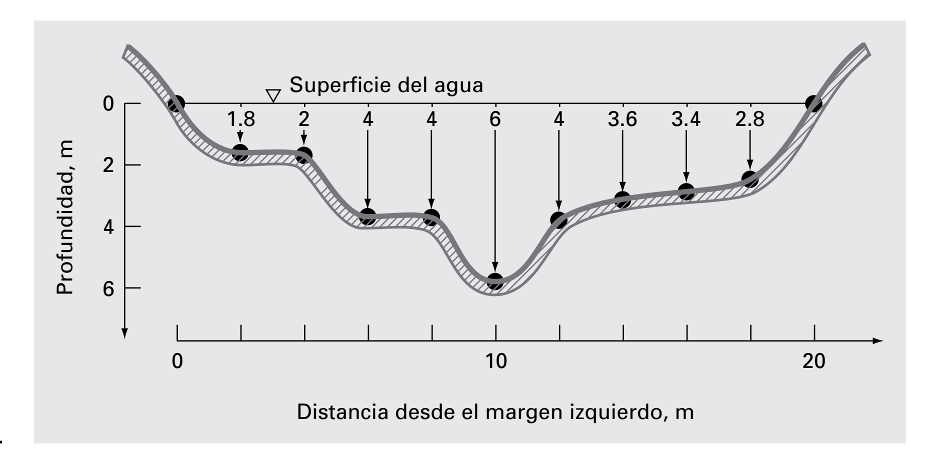


Figura P24.17
Sección transversal de una corriente.

x, m	0	0.375	0.75	1.125	1.5	1.875
y, cm	0	–0.2571	–0.9484	–1.9689	–3.2262	–4.6414

x, m	2.25	2.625	3
y, cm	–6.1503	–7.7051	–9.275

Utilice diferenciación numérica para calcular la pendiente, el momento (en N m), la fuerza cortante (en N) y la carga distribuida (en N/m). Use los siguientes valores de parámetros en su cálculo: E = 200 GPa e I = 0.0003 m^4.

24.21 Un estudio de ingeniería del transporte requiere que se calcule el número total de autos que cruzan por una intersección en un periodo de 24 horas. Un individuo la visita en diferentes momentos durante el curso de un día y cuenta durante un minuto los autos que pasan por la intersección. Utilice los datos que se resumen en la tabla P24.21 para estimar el número total de autos que cruzan por día (tenga cuidado con las unidades).

24.22 Se midió la fuerza del viento distribuida contra el costado de un rascacielos, obteniendo:

Altura, l, m	0	30	60	90	120	150	180
Fuerza, $F(l)$, N/m	0	340	1 200	1 600	2 700	3 100	3 200

Altura, l, m	210	240
Fuerza, $F(l)$, N/m	3 500	3 800

Calcule la fuerza neta y la línea de acción debida a este viento distribuido.

24.23 El agua ejerce presión sobre el parámetro de una presa, como se ilustra en la figura P24.23. La presión se describe con la ecuación:

$$p(z) = \rho g(D - z) \tag{P24.23.1}$$

donde $p(z)$ es la presión en Pascales (o N/m^2) que se ejerce a z metros de elevación sobre el fondo de la presa; ρ = densidad del agua, que para este problema se supone ser constante de 10^3 kg/m^3; g = aceleración de la gravedad (9.8 m/s^2), y D = elevación (en m) del agua desde el fondo de la presa hasta la superficie. De acuerdo con la ecuación (P24.23.1), la presión se incrementa en forma lineal con la profundidad, como se ilustra en la figura P24.23*a*. Si se omite la presión atmosférica (porque opera contra ambos lados de la cara de la presa y en esencia se cancela), la fuerza total f_t se determina con la multiplicación de la presión por el área de la cara de la presa (como se muestra en la figura P24.23*b*). Como tanto la presión como el área varían con la elevación, la fuerza total se obtiene con la evaluación de:

$$f_t = \int_0^D \rho g w(z)(D - z)\,dz$$

donde $w(z)$ = ancho de la cara de la presa (m) en la elevación z (véase la figura P24.23*b*). La línea de acción también puede obtenerse con la evaluación de:

$$d = \frac{\int_0^D \rho g z w(z)(D - z)\,dz}{\int_0^D \rho g w(z)(D - z)\,dz}$$

Use la regla de Simpson para calcular f_t y d. Compruebe los resultados con su programa de cómputo para la regla del trapecio.

24.24 Para estimar el tamaño de una presa nueva, usted tiene que determinar el volumen total de agua (m^3) que fluye por un río en un año. Usted dispone de los datos históricos promedio para el río:

Fecha	Med Ene	Med Feb	Med Mar	Med Abr	Med Jun	Med Sep	Med Oct	Med Nov	Med Dic
Flujo, m^3/s	30	38	82	125	95	20	22	24	35

Determine el volumen. Tenga cuidado con las unidades y al hacer una estimación apropiada del flujo en los puntos extremos.

24.25 Los datos que se enlistan en la tabla siguiente proporcionan mediciones por hora del flujo de calor q (cal/cm^2/h) en la superficie de un colector solar. Como ingeniero arquitecto, usted debe estimar

Tabla P24.21 Tasa de flujo de tráfico (autos/min) en una intersección medida en diferentes momentos durante un periodo de 24 horas.

Hora	Tasa	Hora	Tasa	Hora	Tasa
12:00 medianoche	2	9:00 a.m.	11	6:00 p.m.	20
2:00 a.m.	2	10:30 a.m.	4	7:00 p.m.	10
4:00 a.m.	0	11:30 a.m.	11	8:00 p.m.	8
5:00 a.m.	2	12:30 p.m.	12	9:00 p.m.	10
6:00 a.m.	6	2:00 p.m.	8	10:00 p.m.	8
7:00 a.m.	7	4:00 p.m.	7	11:00 p.m.	7
8:00 a.m.	23	5:00 p.m.	26	12:00 medianoche	3

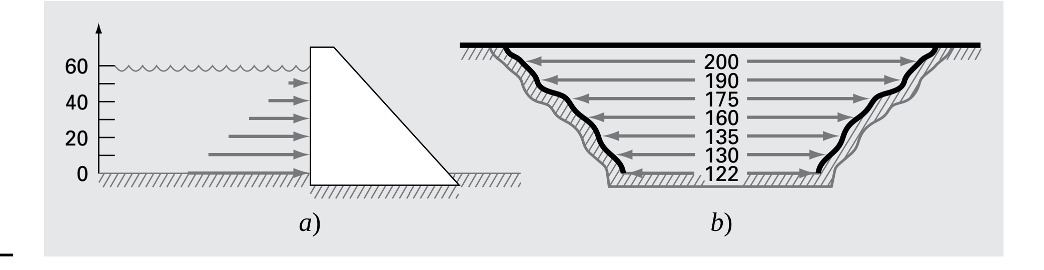


Figura P24.23
Presión que ejerce el agua sobre el parámetro de una presa: *a*) vista lateral que muestra el incremento lineal de la fuerza con la profundidad; *b*) vista frontal donde se muestra el ancho de la presa en metros.

el calor total absorbido por un panel colector de 150 000 cm^2 durante un periodo de 14 horas. El panel tiene una eficiencia de absorción e_{ab} de 45%. El calor total absorbido está dado por:

$$h = e_{ab}\int_0^t qA\,dt$$

donde A es el área y q el flujo de calor.

t	0	2	4	6	8	10	12	14
q	0.10	5.32	7.80	8.00	8.03	6.27	3.54	0.20

24.26 El flujo de calor q es la cantidad de calor que fluye a través de una unidad de área de cierto material por unidad de tiempo. Se calcula con la ley de Fourier:

$$J = -k\frac{dT}{dx}$$

donde J está en unidades de J/m^2/s o W/m^2, y k es un coeficiente de la conductividad térmica que parametriza las propiedades conductoras de calor del material y se expresa en unidades de W/(°C · m). T = temperatura (°C); y x = distancia (m) a lo largo de la trayectoria del flujo de calor. La ley de Fourier la emplean en forma rutinaria los ingenieros arquitectos para determinar el flujo de calor a través de las paredes. Se midieron las temperaturas siguientes a partir de la superficie (x = 0) de una pared de piedra:

x, m	0	0.08	0.16
T, °C	20	17	15

Si el flujo en x = 0 es de 60 W/m^2, calcule el valor de k.

24.27 El área de la superficie horizontal A_s (m^2) de un lago, a cierta profundidad, se calcula a partir del volumen por medio de diferenciación:

$$A_s(z) = -\frac{dV}{dz}(z)$$

donde V = volumen (m^3) y z = profundidad (m), se mide a partir de la superficie en dirección del fondo. La concentración promedio de una sustancia que varía con la profundidad $\bar{c}$ (g/m^3) se obtiene por integración:

$$\bar{c} = \frac{\int_0^Z c(z)A_s(z)\,dz}{\int_0^Z A_s(z)\,dz}$$

donde Z = profundidad total (m). Determine la concentración promedio con base en los datos siguientes:

z, m	0	4	8	12	16
V, 10^6 m^3	9.8175	5.1051	1.9635	0.3927	0.0000
c, g/m^3	10.2	8.5	7.4	5.2	4.1

Ingeniería eléctrica

24.28 Lleve a cabo el mismo cálculo que en la sección 24.3, pero para la corriente según las especificaciones siguientes:

$$i(t) = 5e^{-1.25t}\text{ sen } 2\pi\, t \qquad \text{para } 0 \le t \le T/2$$
$$i(t) = 0 \qquad \text{para } T/2 < t \le T$$

donde T = 1 s. Use la cuadratura de Gauss de cinco puntos para estimar la integral.

24.29 Repita el problema 24.28, pero emplee la regla de Simpson 1/3 de cinco segmentos.

24.30 Vuelva a hacer el problema 24.28, pero use integración de Romberg con ε_s = 1%.

24.31 La ley de Faraday caracteriza la caída de voltaje a través de un inductor, así:

$$V_L = L\frac{di}{dt}$$

donde V_L = caída del voltaje (V), L = inductancia (en henrios; 1 H = 1 V · s/A), i = corriente (A) y t = tiempo (s). Determine la caída del voltaje como función del tiempo, con los datos siguientes para una inductancia de 4 H.

t	0	0.1	0.2	0.3	0.5	0.7
i	0	0.16	0.32	0.56	0.84	2.0

24.32 Con base en la ley de Faraday (véase el problema 24.31), use los datos siguientes de voltaje para estimar la inductancia en henrios si se pasa durante 400 milisegundos una corriente de 2 A por el inductor.

t, ms	0	10	20	40	60	80	120	180	280	400
V, volts	0	18	29	44	49	46	35	26	15	7

24.33 Suponga que la corriente a través de una resistencia está descrita por la función:

$$i(t) = (60 - t)^2 + (60 - t)\text{ sen }(\sqrt{t})$$

y que la resistencia es función de la corriente:

$$R = 10i + 2i^{2/3}$$

Calcule el voltaje promedio desde t = 0 hasta 60, con el uso de la regla de Simpson 1/3 de segmentos múltiples.

24.34 Si inicialmente un capacitor no tiene carga, el voltaje a través de él como función del tiempo se calcula por medio de:

$$V(t) = \frac{1}{C}\int_0^t i(t)\,dt$$

Si $C = 10^{-5}$ faradios, use los datos de corriente que siguen para elaborar una gráfica del voltaje *versus* el tiempo:

t, s	0	0.2	0.4	0.6	0.8	1	1.2
i, 10^{-3} A	0.2	0.3683	0.3819	0.2282	0.0486	0.0082	0.1441

Ingeniería mecánica/aeroespacial

24.35 Realice el mismo cálculo que en la sección 24.4, pero use las siguientes ecuaciones:

$$F(x) = 1.6x - 0.045x^2$$
$$\theta(x) = 0.8 + 0.125x - 0.009x^2 + 0.0002x^3$$

Use reglas trapezoidales de 4, 8 y 16 segmentos para calcular la integral.

24.36 Repita el problema 24.35, pero use *a*) la regla de Simpson 1/3, *b*) la integración de Romberg hasta ε_s = 0.5% y *c*) la cuadratura de Gauss.

24.37 Calcule el trabajo como se describe en la sección 24.4, pero use las siguientes ecuaciones para $F(x)$ y $\theta(x)$:

$$F(x) = 1.6x - 0.045x^2$$

$$\theta(x) = -0.00055x^3 + 0.0123x^2 + 0.13x$$

La fuerza está en newtons y el ángulo en radianes. Realice la integración desde $x = 0$ hasta 30 m.

24.38 Como se hizo en la sección 24.4, determine el trabajo realizado si una fuerza constante de 1 N aplicada a un ángulo θ resulta en los siguientes desplazamientos. Use la función de MATLAB `cumtrapz` para determinar el trabajo acumulativo y grafique el resultado contra θ.

x, m	0	1	2.7	3.8	3.7	3	1.4
θ, grados	0	30	60	90	120	150	180

24.39 El trabajo que realiza un objeto es igual a la fuerza por la distancia que se desplaza en la dirección de la fuerza. La velocidad de un objeto en la dirección de una fuerza está dada por

$$v = 4t \qquad 0 \le t \le 4$$
$$v = 16 + (4 - t)^2 \qquad 4 \le t \le 14$$

donde v = m/s. Emplee la aplicación múltiple de la regla de Simpson para determinar el trabajo si se aplica una fuerza constante de 200 N para toda t.

24.40 La tasa de enfriamiento de un cuerpo (figura P24.40) se expresa como:

$$\frac{dT}{dt} = -k(T - T_a)$$

donde T = temperatura del cuerpo (°C), T_a = temperatura del medio circundante (°C) y k = constante de proporcionalidad (por minuto). Así, esta ecuación (denominada *ley de Newton para el enfriamiento*) especifica que la tasa de enfriamiento es proporcional a la diferencia de temperaturas del cuerpo y del medio circundante. Si una bola de metal calentada a 80°C se sumerge en agua que se mantiene a T_a = 20°C constante, la temperatura de la bola cambia, así

Tiempo, min	0	5	10	15	20	25
T, °C	80	44.5	30.0	24.1	21.7	20.7

Utilice diferenciación numérica para determinar dT/dt en cada valor del tiempo. Grafique dT/dt *versus* $T - T_a$, y emplee regresión lineal para evaluar k.

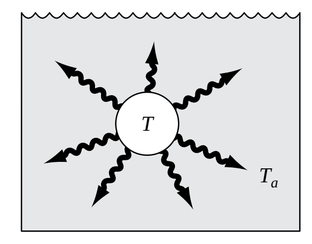


Figura P24.40

24.41 Una barra sujeta a una carga axial (véase la figura P24.41*a*) se deformará como se ilustra en la curva esfuerzo-tensión que aparece en la figura P24.41*b*. El área bajo la curva desde el esfuerzo cero hasta el punto de ruptura se denomina *módulo de rigidez* del material. Proporciona una medida de la energía por unidad de volumen que se requiere para hacer que el material se rompa. Por ello, es representativo de la capacidad del material para superar una carga de impacto. Use integración numérica para calcular el módulo de rigidez para la curva esfuerzo-tensión que se aprecia en la figura P24.41*b*.

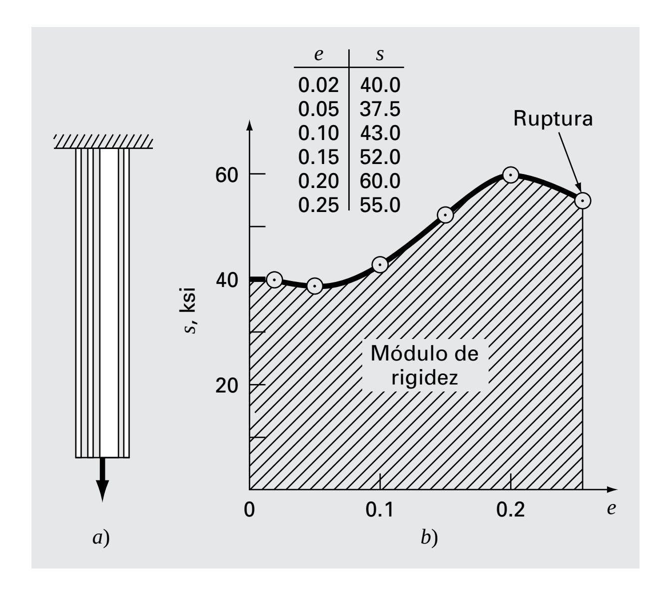


Figura P.24.41
a) Barra sujeta a carga axial y *b*) la curva resultante esfuerzo-tensión, en la que el esfuerzo está en kips por pulgada cuadrada (10^3 lb/pulg^2), y la tensión es adimensional.

24.42 Si se conoce la distribución de la velocidad de un fluido a través de un tubo (véase la figura P24.42), la tasa de flujo Q (es decir, el volumen de agua que pasa por el tubo por unidad de tiempo) se calcula por medio de $Q = \int v dA$, donde v es la velocidad y A es el área de la sección transversal del tubo. (Para entender el significado físico de esta relación, recuerde la estrecha conexión que hay entre la suma y la integración.) Para un tubo circular, $A = \pi r^2$ y $dA = 2\pi r\, dr$. Por lo tanto,

$$Q = \int_0^r v(2\pi r)\, dr$$

donde r es la distancia radial medida hacia fuera del centro del tubo. Si la distribución de la velocidad está dada por

$$v = 2\left(1 - \frac{r}{r_0}\right)^{1/6}$$

donde r_0 es el radio total (en este caso, 3 cm), calcule Q con el empleo de la regla del trapecio de aplicación múltiple. Analice los resultados.

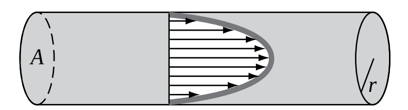


Figura P24.42

24.43 Con los datos siguientes, calcule el trabajo realizado por el estiramiento de un resorte con una constante de estiramiento de k = 300 N/m a x = 0.35 m.

F, 10^3N	0	0.01	0.028	0.046	0.063	0.082	0.11	0.13
x, m	0	0.05	0.10	0.15	0.20	0.25	0.30	0.35

24.44 Se midió la posición de un avión de combate durante su aterrizaje en la cubierta de un portaviones:

t, s	0	0.52	1.04	1.75	2.37	3.25	3.83
x, m	153	185	208	249	261	271	273

donde x es la distancia desde el extremo del portaviones. Estime *a*) la velocidad (dx/dt) y *b*) la aceleración (dv/dt), por medio de diferenciación numérica.

24.45 Emplee la regla de Simpson de aplicación múltiple para evaluar la distancia vertical que recorre un cohete si su velocidad vertical está dada por:

$v = 11t^2 - 5t \qquad 0 \le t \le 10$

$v = 1\,100 - 5t \qquad 10 \le t \le 20$

$v = 50t + 2(t-20)^2 \qquad 20 \le t \le 30$

Además, use diferenciación numérica para desarrollar gráficas de la aceleración (dv/dt) y el arranque (d^2v/dt^2) contra el tiempo, para t = 0 a 30. Observe que el arranque es muy importante porque está altamente relacionado con lesiones como traumatismo cervical.

24.46 La velocidad hacia arriba de un cohete se calcula con la fórmula que sigue:

$$v = u \ln\left(\frac{m_0}{m_0 - qt}\right) - gt$$

donde v = velocidad hacia arriba, u = velocidad a que se expele el combustible en relación con el cohete, m_0 = masa inicial del cohete en el tiempo $t = 0$, q = tasa de consumo de combustible y g = aceleración de la gravedad en dirección hacia abajo (se supone constante = 9.8 m/s^2). Si u = 1 800 m/s, m_0 = 160 000 kg y q = 2 500 kg/s, utilice la regla del trapecio de seis segmentos y de Simpson 1/3, la cuadratura de Gauss de seis puntos y los métodos de Romberg $O(h^8)$ para determinar qué altura alcanzará el cohete en un vuelo de 30 s. Además, use diferenciación numérica para generar una gráfica de aceleración como función del tiempo.

24.47 En relación con los datos del problema 20.61, encuentre la tasa de tensión por medio de métodos de diferencias finitas. Use aproximaciones de la derivada con exactitud de segundo orden y grafique sus resultados. Al ver la gráfica, es evidente que existe algún error experimental de inicio. Encuentre la media y desviación estándar de la tasa de tensión después de eliminar los puntos de datos que representen el error experimental de arranque.

24.48 Un flujo desarrollado por completo que pasa a través de un tubo de 40 cm de diámetro tiene el perfil de velocidad siguiente:

Radio, *r*, cm	0.0	2.5	5.0	7.5	10.0	12.5
Velocidad, *v*, m/s	0.914	0.890	0.847	0.795	0.719	0.543

Radio, *r*, cm	15.0	17.5	20.0
Velocidad, *v*, m/s	0.427	0.204	0

Encuentre la tasa de flujo volumétrico, Q, con la relación $Q = \int_0^R 2\pi rv\,dr$, donde r es el eje radial del tubo, R es el radio del tubo, y v es la velocidad. Resuelva el problema con dos enfoques diferentes.

a) Ajuste una curva polinomial a los datos de velocidad e intégrela en forma analítica.
b) Para la integración utilice una aplicación múltiple de la regla de Simpson 1/3.
c) Encuentre el error porcentual con el uso de la integral del ajuste polinomial como el valor más correcto.

24.49 Un fluido desarrollado por completo de un plástico de Bingham que se mueve por un tubo de 12 pulg de diámetro, tiene el perfil de velocidades que sigue. El flujo de un fluido de Bingham no corta el núcleo central, lo que produce un flujo tapón alrededor de la línea central.

Radio, *r*, pulg	0	1	2	3	4	5	6
Velocidad, *v*, pie/s	5.00	5.00	4.62	4.01	3.42	1.69	0.00

Encuentre la tasa de flujo volumétrico total, Q, con el uso de la relación $Q = \int_{r_1}^{r_2} 2\pi rv\,dr + v_c A_c$, donde r es el eje radial del tubo, v es la velocidad, v_c es la velocidad en el núcleo y A_c es el área de la sección transversal del tapón. Resuelva el problema con dos enfoques distintos.

a) Ajuste una curva polinomial a los datos fuera del núcleo e intégrela.
b) Para la integración emplee la regla de Simpson de aplicaciones múltiples.
c) Encuentre el error porcentual con el uso de la integral del ajuste polinomial como el valor más correcto.

24.50 La entalpía de un gas real es función de la presión como se describe a continuación. Los datos se tomaron para un fluido real. Estime la entalpía del fluido a 400 K y 50 atm (evalúe la integral de 0.1 atm a 50 atm).

$$H \int_0^P \left(V - T\left(\frac{\partial V}{\partial T}\right)_P\right) dP$$

	***V*, L**		
***P*, atm**	***T* = 350 K**	***T* = 400 K**	***T* = 450 K**
0.1	220	250	282.5
5	4.1	4.7	5.23
10	2.2	2.5	2.7
20	1.35	1.49	1.55
25	1.1	1.2	1.24
30	0.90	0.99	1.03
40	0.68	0.75	0.78
45	0.61	0.675	0.7
50	0.54	0.6	0.62

24.51 Dados los datos siguientes, encuentre el trabajo isotérmico realizado sobre el gas cuando se comprime de 23 L a 3 L (recuerde que $W = -\int_{V_1}^{V_2} P\,dV$).

V, L	3	8	13	18	23
P, atm	12.5	3.5	1.8	1.4	1.2

a) Encuentre en forma numérica el trabajo realizado sobre el gas, con la regla del trapecio de 1, 2 y 4 segmentos.
b) Calcule las razones de los errores en estas estimaciones y relaciónelas con el análisis del error de la regla del trapecio de multiplicación del capítulo 21.

24.52 La ecuación de Rosin-Rammler-Bennet (RRB) se emplea para describir la distribución de los tamaños del polvo fino. $F(x)$ representa la masa acumulada de las partículas de polvo de diámetro x y más pequeñas. x' y n' son constantes iguales a 30 μm y 1.44, respectivamente. La distribución de la densidad de masa $f(x)$ o masa de las partículas de polvo de un diámetro x, se encuentra con la derivada de la distribución acumulada.

$$F(x) = 1 - e^{-(x/x')^{n'}} \qquad f(x) = \frac{dF(x)}{dx}$$

a) Calcule en forma numérica la distribución de la densidad de masa *f*(*x*), y grafique tanto *f*(*x*) como la distribución acumulada *F*(*x*).

b) Con sus resultados del inciso *a*), calcule la moda del tamaño de la distribución de la densidad de masa —es decir, el tamaño en que la derivada de *f*(*x*) es igual a cero—.

c) Encuentre el área superficial por masa de polvo S_m (cm^2/g), por medio de:

$$S_m = \frac{6}{\rho}\int_{d_{\text{mín}}}^{\infty} \frac{f(x)}{x}\, dx$$

La ecuación es válida sólo para partículas esféricas. Suponga una densidad $\rho = 1$ g cm^{-3} y un diámetro mínimo, $d_{\text{mín}}$, de polvo incluido en la distribución, de 1 μm.

24.53 Para el flujo de un fluido sobre una superficie, el flujo de calor hacia la superficie se calcula con:

$$J = -k\frac{dT}{dy}$$

donde J = flujo de calor (W/m^2), k = conductividad térmica (W/m · K), T = temperatura (K) y y = distancia normal a la superficie (m). Se hicieron las mediciones siguientes para el flujo de aire sobre una placa plana que mide 200 cm de largo y 50 cm de ancho:

y, cm	0	1	3	5
T, K	900	480	270	200

Si k = 0.028 J/s · m · K, *a*) determine el flujo a la superficie, y *b*) la transferencia de calor en watts. Observe que 1 J = 1 W · s.

24.54 El gradiente de presión para un flujo laminar a través de un tubo de radio constante, está dado por:

$$\frac{dp}{dx} = -\frac{8\mu Q}{\pi r^4}$$

donde p = presión (N/m^2), x = distancia a lo largo de la línea central del tubo (m), μ = viscosidad dinámica (N · s/m^2), Q = flujo (m^3/s) y r = radio (m).

a) Determine la caída de presión para un tubo de 10 cm de longitud para un líquido viscoso (μ = 0.005 N · s/m^2, densidad = $\rho = 1 \times 10^3$ kg/m^3) con un flujo de 10×10^{-6} m^3/s, y las variaciones del radio con la longitud que siguen,

x, cm	0	2	4	5	6	7	10
r, mm	2	1.35	1.34	1.6	1.58	1.42	2

b) Compare su resultado con la caída de presión que tendría que ocurrir si el tubo tuviera un radio constante igual al radio promedio.

c) Determine el número de Reynolds promedio para el tubo a fin de comprobar que el flujo es de verdad laminar ($Re = \rho v D/\mu < 2\,100$, donde v = velocidad).

24.55 Se recabaron datos de la velocidad del aire en radios diferentes desde la línea central de un tubo circular de 16 cm de diámetro, como se muestra a continuación:

r, cm	0	1.60	3.20	4.80	6.40	7.47	7.87	7.95	8
v, m/s	10	9.69	9.30	8.77	7.95	6.79	5.57	4.89	0

Utilice integración numérica para determinar la tasa de flujo de masa, que se calcula como:

$$\int_0^R \rho v 2\pi r\ dr$$

donde ρ = densidad (= 1.2 kg/m^3). Exprese sus resultados en kg/s.

EPÍLOGO: PARTE SEIS

PT6.4 ALTERNATIVAS

La tabla PT6.4 ofrece un resumen de las ventajas y las desventajas en la integración o cuadratura numérica. La mayoría de esos métodos se basa en la interpretación geométrica de considerar una integral como el área bajo una curva. Estas técnicas están diseñadas para evaluar la integral en dos casos diferentes: 1) una función matemática y 2) datos discretos en forma tabular.

Las fórmulas de Newton-Cotes son los principales métodos analizados en el capítulo 21. Se aplican a funciones, continuas y discretas, existen versiones tanto cerradas como abiertas. Las fórmulas abiertas tienen límites de integración que se extienden más allá del intervalo donde aparecen los datos, muy rara vez se utilizan para la evaluación de integrales definidas. Sin embargo, son de utilidad para la solución de ecuaciones diferenciales ordinarias y para la evaluación de integrales impropias.

Las fórmulas cerradas de Newton-Cotes se basan en el reemplazo de una función matemática o de datos tabulados, por un polinomio de interpolación que es fácil de integrar. La versión más simple es la regla del trapecio, que se basa en el cálculo del área bajo una línea recta que une valores adyacentes de la función. Una manera para aumentar la exactitud de la regla del trapecio consiste en subdividir el intervalo de integración, desde a hasta b, en varios segmentos y aplicar el método a cada uno de ellos.

Además de aplicar la regla del trapecio con una segmentación más fina, otra forma de obtener una estimación más exacta de la integral es usar polinomios de mayor grado que unan los puntos. Si se emplea una ecuación cuadrática, el resultado es la regla de Simpson 1/3; si se utiliza una ecuación cúbica, será la regla de Simpson 3/8. Como estas fórmulas son mucho más exactas que la regla del trapecio, por lo común tienen mayor preferencia, disponiéndose de versiones de aplicación múltiple. En situaciones con un número de segmentos par, se recomienda la aplicación múltiple de la regla de Simpson 1/3. Para un número de segmentos impar se puede aplicar la regla de Simpson 3/8 a los últimos tres segmentos, y la regla de Simpson 1/3 a los segmentos restantes.

También existen fórmulas de Newton-Cotes de mayor grado. Sin embargo, en la práctica se usan muy rara vez. Cuando se requiere de alta exactitud, se dispone de las fórmulas de integración de Romberg y de la cuadratura de Gauss. Debe observarse que ambas tienen mayor valor práctico en los casos donde se dispone de la función. Dichas técnicas no son tan adecuadas para datos tabulados.

PT6.5 RELACIONES Y FÓRMULAS IMPORTANTES

La tabla PT6.5 resume la información importante que se expuso en la parte seis. Esta tabla se puede consultar para un acceso rápido de las relaciones y fórmulas importantes.

Tabla PT6.4 Comparación de las características de los distintos métodos para la integración numérica. Las comparaciones se basan en la experiencia general y no toman en cuenta el comportamiento de funciones especiales.

Método	Puntos necesarios para una aplicación	Puntos requeridos para n aplicaciones	Error de truncamiento	Aplicación	Dificultad de programación	Comentarios
Regla del trapecio	2	$n + 1$	$\simeq h^3 f''(\xi)$	Amplia	Fácil	
Regla de Simpson 1/3	3	$2n + 1$	$\simeq h^5 f^{(4)}(\xi)$	Amplia	Fácil	
Regla de Simpson 1/3 y 3/8	3 o 4	≥ 3	$\simeq h^5 f^{(4)}(\xi)$	Amplia	Fácil	
Newton-Cotes de mayor grado	≥ 5	N/D	$\simeq h^7 f^{(6)}(\xi)$	Rara	Fácil	
Integración de Romberg	3			Requiere que se conozca $f(x)$	Moderada	No es tan apropiado para datos tabulares
Cuadratura de Gauss	≥ 2	N/D		Requiere que se conozca $f(x)$	Fácil	No es tan apropiado para datos tabulares

PT6.6 MÉTODOS AVANZADOS Y REFERENCIAS ADICIONALES

Aunque revisamos varias de las técnicas de integración numérica, hay otros métodos que tienen utilidad en la práctica de la ingeniería. Por ejemplo, los esquemas adaptativos para resolver ecuaciones diferenciales ordinarias, permiten evaluar integrales complicadas, como se analizará en el capítulo 25.

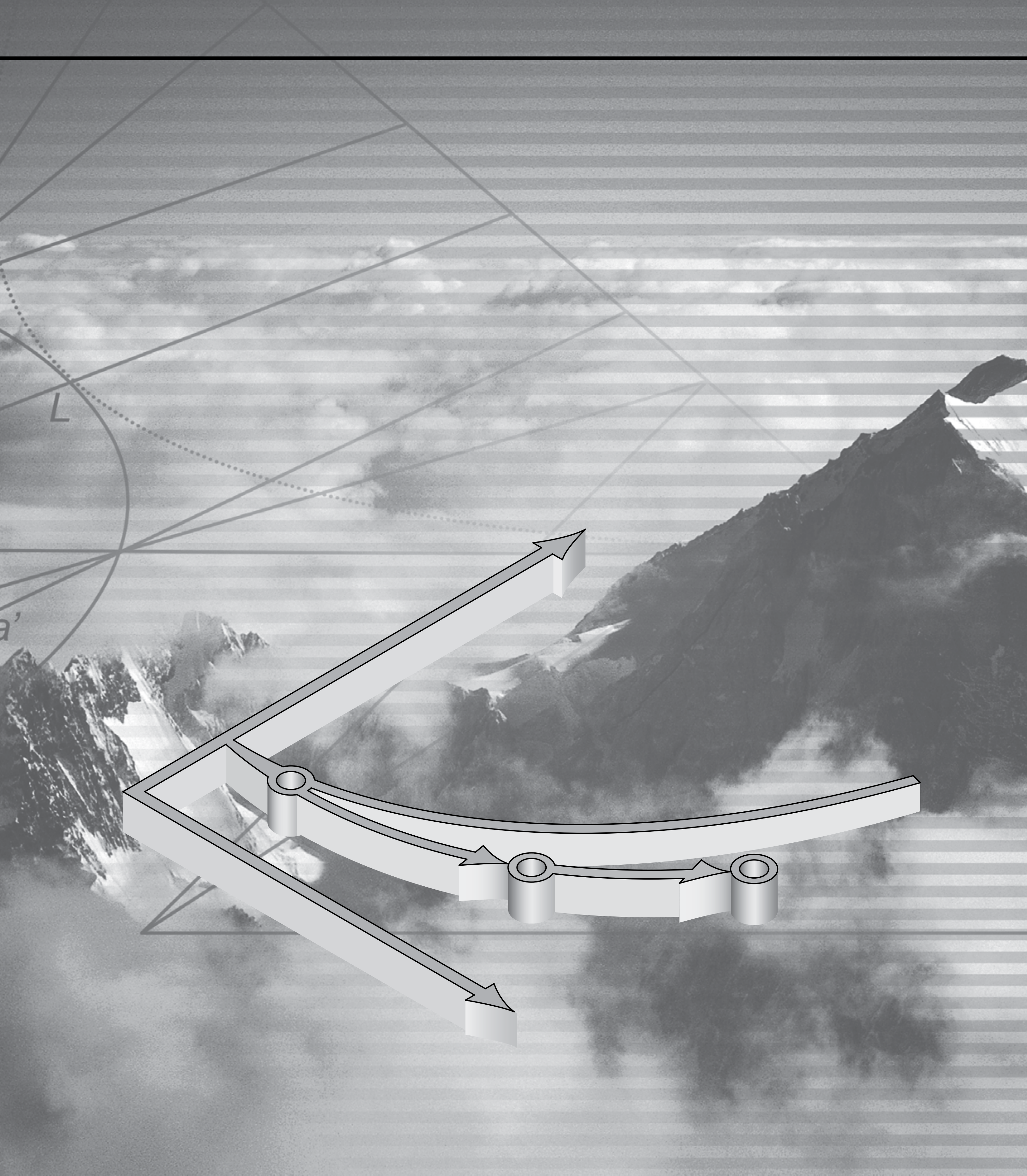
Otro método para calcular integrales consiste en ajustar splines cúbicos a los datos. Las ecuaciones cúbicas resultantes se integran de manera fácil (Forsythe y cols., 1977). Algunas veces se usa un procedimiento similar también en diferenciación. Por último, además de las fórmulas de Gauss-Legendre analizadas en la sección 22.3, existen otras fórmulas de cuadratura. Carnahan, Luther y Wilkes (1969), y Ralston y Rabinowitz (1978), resumen muchos de esos procedimientos.

En síntesis, lo anterior tiene la intención de proporcionarle algunos caminos para una exploración más profunda de este tema. Además, todas las referencias anteriores describen las técnicas básicas tratadas en la parte seis. Le recomendamos consultar estas fuentes alternativas para ampliar su conocimiento de los métodos numéricos para la integración.

Tabla PT6.5 Resumen de fórmulas importantes presentadas en la parte seis.

Método	Formulación	Interpretación gráfica	Error
Regla del trapecio	$I \simeq (b-a)\dfrac{f(a)+f(b)}{2}$	$f(x)$; a, b, x	$-\dfrac{(b-a)^3}{12}f''(\xi)$
Regla del trapecio de aplicación múltiple	$I \simeq (b-a)\dfrac{f(x_0)+2\sum_{i=1}^{n-1}f(x_i)+f(x_n)}{2n}$	$f(x)$; $a = x_0$, $b = x_n$, x	$-\dfrac{(b-a)^3}{12n^2}\bar{f}''$
Regla de Simpson 1/3	$I \simeq (b-a)\dfrac{f(x_0)+4f(x_1)+f(x_2)}{6}$	$f(x)$; $a = x_0$, $b = x_2$, x	$-\dfrac{(b-a)^5}{2\,880}f^{(4)}(\xi)$
Regla de Simpson 1/3 de aplicación múltiple	$I \simeq (b-a)\dfrac{f(x_0)+4\sum_{i=1,3}^{n-1}f(x_i)+2\sum_{i=2,4}^{n-2}f(x_i)+f(x_n)}{3n}$	$f(x)$; $a = x_0$, $b = x_n$, x	$-\dfrac{(b-a)^5}{180n^4}f^{(4)}$
Regla de Simpson 3/8	$I \simeq (b-a)\dfrac{f(x_0)+3f(x_1)+3f(x_2)+f(x_3)}{8}$	$f(x)$; $a = x_0$, $b = x_3$, x	$-\dfrac{(b-a)^5}{6\,480}f^{(4)}(\xi)$
Integración de Romberg	$I_{j,k} = \dfrac{4^{k-1}I_{j+1,k-1}-I_{j,k-1}}{4^{k-1}-1}$	$I_{j,k-1} \longrightarrow I_{j,k}$; $I_{j+1,k-1} \nearrow$	$O(h^{2k})$
Cuadratura de Gauss	$I \simeq c_0 f(x_0)+c_1 f(x_1)+\cdots+c_{n-1}f(x_{n-1})$	$f(x)$; x_0, x_1, x	$\simeq f^{(2n+2)}(\xi)$

PARTE SIETE



ECUACIONES DIFERENCIALES ORDINARIAS

PT7.1 MOTIVACIÓN

En el primer capítulo de este libro obtuvimos la siguiente ecuación basada en la segunda ley de Newton, para calcular la velocidad v del paracaidista en caída como una función del tiempo t [recuerde la ecuación (1.9)]:

$$\frac{dv}{dt} = g - \frac{c}{m}v \tag{PT7.1}$$

donde g es la constante gravitacional, m es la masa y c es el coeficiente de arrastre. Tales ecuaciones, que se componen de una función desconocida y de sus derivadas, se conocen como *ecuaciones diferenciales*. A la ecuación (PT7.1) algunas veces se le llama una *ecuación de razón de cambio*, ya que expresa la razón de cambio de una variable como una función de variables y parámetros. Estas ecuaciones desempeñan un papel importante en ingeniería debido a que muchos fenómenos físicos se formulan matemáticamente mejor en términos de su razón de cambio.

En la ecuación (PT7.1), la cantidad que se está derivando, v, se conoce como *variable dependiente*. La cantidad con respecto a la cual v se está derivando, t, se conoce como *variable independiente*. Cuando la función tiene una variable independiente, la ecuación se llama *ecuación diferencial ordinaria* (o *EDO*). Esto contrasta con una *ecuación diferencial parcial* (o *EDP*) que involucra dos o más variables independientes.

Las ecuaciones diferenciales se clasifican también en cuanto a su orden. Por ejemplo, la ecuación (PT7.1) se denomina como *EDO de primer orden*, ya que la derivada mayor es una primera derivada. Una *EDO de segundo orden* tiene una segunda derivada, como la mayor. Por ejemplo, la ecuación que describe la posición x de un sistema masa-resorte con amortiguamiento es la EDO de segundo orden,

$$m\frac{d^2x}{dt^2} + c\frac{dx}{dt} + kx = 0 \tag{PT7.2}$$

donde c es un coeficiente de amortiguamiento y k es la constante del resorte. De manera similar, una ecuación de n-ésimo orden tiene una n-ésima derivada, como la mayor.

Las ecuaciones de orden superior pueden reducirse a un sistema de ecuaciones de primer orden. Para la ecuación (PT7.2), esto se logra al definir una nueva variable y, donde

$$y = \frac{dx}{dt} \tag{PT7.3}$$

que al derivar con respecto a t se obtiene

$$\frac{dy}{dt} = \frac{d^2x}{dt^2} \tag{PT7.4}$$

Las ecuaciones (PT7.3) y (PT7.4) se sustituyen después en la ecuación (PT7.2) para llegar a

$$m\frac{dy}{dt} + cy + kx = 0 \tag{PT7.5}$$

o

$$\frac{dy}{dt} = -\frac{cy + kx}{m} \tag{PT7.6}$$

Así, las ecuaciones (PT7.3) y (PT7.6) son un sistema de ecuaciones diferenciales de primer orden, equivalentes a la ecuación de segundo orden original. Como otras ecuaciones diferenciales de *n*-ésimo orden pueden reducirse en forma similar, esta parte de nuestro libro se concentra en la solución de ecuaciones diferenciales de primer orden. Algunas aplicaciones de la ingeniería en el capítulo 28 tratan con la solución de EDO de segundo orden por reducción a un sistema de dos ecuaciones diferenciales de primer orden.

PT7.1.1 Métodos para resolver EDO sin el uso de la computadora

Sin una computadora, las EDO podrían resolverse usando técnicas de integración analítica. Por ejemplo, la ecuación (PT7.1) se multiplica por dt y se integra para obtener

$$v = \int \left(g - \frac{c}{m} v \right) dt \tag{PT7.7}$$

El lado derecho de esta ecuación se conoce como *integral indefinida* debido a que no se especifican los límites de integración. Esto contrasta con las integrales definidas que se analizaron en la parte seis [compare la ecuación (PT7.7) con la ecuación (PT6.6)].

Una solución analítica para la ecuación (PT7.7) se obtiene si la integral indefinida puede evaluarse en forma exacta como una ecuación. Por ejemplo, recuerde que para el problema del paracaidista en caída, la ecuación (PT7.7) se resolvió analíticamente con la ecuación (1.10) (suponga que $v = 0$ en $t = 0$):

$$v(t) = \frac{gm}{c}\left(1 - e^{-(c/m)t}\right) \tag{1.10}$$

La mecánica para obtener tales soluciones analíticas se analizará en la sección PT7.2. Mientras tanto, lo relevante es que no están disponibles las soluciones exactas para muchas EDO de importancia práctica. Como sucede en la mayoría de las situaciones analizadas en otras partes de este libro, los métodos numéricos ofrecen la única alternativa viable para tales casos. Como estos métodos numéricos por lo común requieren de computadoras, antes del auge de la informática los ingenieros se veían muy limitados en el alcance de sus investigaciones.

Un método muy importante que los ingenieros y los matemáticos desarrollaron para superar este dilema fue la *linealización*. Una ecuación diferencial ordinaria lineal es aquella que tiene la forma general

$$a_n(x)y^{(n)} + \cdots + a_1(x)y' + a_0(x)y = f(x) \tag{PT7.8}$$

donde $y^{(n)}$ es la n-ésima derivada de y con respecto a x, y las a y f son funciones en términos de x. Esta ecuación se conoce como *lineal* debido a que no hay productos o funciones no lineales de la variable dependiente y y no existen productos de sus derivadas. La importancia práctica de las EDO lineales es que se resuelven analíticamente. En cambio, la mayoría de las ecuaciones no lineales no pueden resolverse de manera exacta. Así, en la era anterior a la computadora, una táctica para resolver ecuaciones no lineales era linealizarlas.

Un ejemplo simple es la aplicación de las EDO para predecir el movimiento de un péndulo oscilante (figura PT7.1). De manera similar como se hizo en el desarrollo del problema del paracaidista en caída, se utiliza la segunda ley de Newton para obtener la siguiente ecuación diferencial (véase la sección 28.4 para más detalles):

$$\frac{d^2\theta}{dt^2} + \frac{g}{l} \operatorname{sen} \theta = 0 \tag{PT7.9}$$

donde θ es el ángulo de desplazamiento del péndulo, g es la constante gravitacional y l es la longitud del péndulo. Esta ecuación no es lineal debido al término sen θ. Una forma de obtener la solución analítica es darse cuenta de que para pequeños desplazamientos del péndulo a partir de su condición de equilibrio (es decir, para valores pequeños de θ),

$$\operatorname{sen} \theta \cong \theta \tag{PT7.10}$$

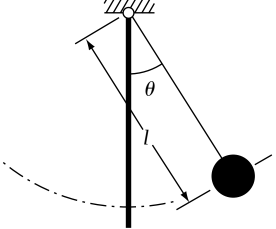


Figura PT7.1
El péndulo oscilante.

Así, si suponemos que nos interesamos sólo en casos donde θ es pequeño, la ecuación (PT7.10) se sustituye en la ecuación (PT7.9) para obtener

$$\frac{d^2\theta}{dt^2}+\frac{g}{l}\theta=0 \tag{PT7.11}$$

Tenemos, por lo tanto, transformada la ecuación (PT7.9) en una forma lineal que es fácil de resolver de manera analítica.

Aunque la linealización es una herramienta muy valiosa para resolver problemas en ingeniería, existen casos donde no se puede utilizar. Por ejemplo, suponga que nos interesa estudiar el comportamiento del péndulo con grandes desplazamientos desde el equilibrio. En tales casos, los métodos numéricos ofrecen una opción viable para obtener la solución. En la actualidad, la disponibilidad tan amplia de las computadoras coloca esta opción al alcance de todos los ingenieros.

PT7.1.2 Las EDO y la práctica en ingeniería

Las leyes fundamentales de la física: la mecánica, la electricidad y la termodinámica con frecuencia se basan en observaciones empíricas que explican las variaciones de las propiedades físicas y los estados de los sistemas. Más que en describir directamente el estado de los sistemas físicos, las leyes a menudo se expresan en términos de los cambios del espacio y del tiempo.

En la tabla PT7.1 se muestran algunos ejemplos. Esas leyes definen mecanismos de cambio. Cuando se combinan con las leyes de conservación de la energía, masa o *momentum*, resultan ecuaciones diferenciales. La integración subsecuente de estas ecuaciones diferenciales origina funciones matemáticas que describen el estado espacial y temporal de un sistema en términos de variaciones de energía, masa o velocidad.

El problema del paracaidista en caída que se presentó en el capítulo 1 es un ejemplo de la obtención de una ecuación diferencial ordinaria, a partir de una ley fundamental. Recuerde que se utilizó la segunda ley de Newton para desarrollar una EDO que describe la razón de cambio de la velocidad de un paracaidista en caída. Al integrar esta expresión, obtenemos una ecuación para predecir la velocidad de caída como una función del tiempo (figura PT7.2). Esta ecuación se utiliza de diferentes formas, entre ellas para propósitos de diseño.

De hecho, tales relaciones matemáticas son la base para la solución de un gran número de problemas de ingeniería. Sin embargo, como se describió en la sección anterior, muchas de las ecuaciones diferenciales de importancia práctica no se pueden resolver utilizando los métodos analíticos de cálculo. Así los métodos que se estudiarán en los siguientes capítulos resultan extremadamente importantes en todos los campos de la ingeniería.

Tabla PT7.1 Ejemplos de las leyes fundamentales que se escriben en términos de la razón de cambio de variables (t = tiempo y x = posición).

Ley	Expresión matemática	Variables y parámetros
Segunda ley de Newton del movimiento	$\frac{dv}{dt}=\frac{F}{m}$	Velocidad (v), fuerza (F) y masa (m)
Ley del calor de Fourier	$q=-k'\frac{dT}{dx}$	Flujo de calor (q), conductividad térmica (k') y temperatura (T)
Ley de difusión de Fick	$J=-D\frac{dc}{dx}$	Flujo másico (J), coeficiente de difusión (D) y concentración (c)
Ley de Faraday (caída de voltaje a través de un inductor)	$\Delta V_L=L\frac{di}{dt}$	Caída de voltaje (ΔV_L), inductancia (L) y corriente (i)

PT7.2 ANTECEDENTES MATEMÁTICOS

La solución de una ecuación diferencial ordinaria es una función en términos de la variable independiente y de parámetros que satisfacen la ecuación diferencial original. Para ilustrar este concepto, empecemos con una función dada

$$y=-0.5x^4+4x^3-10x^2+8.5x+1 \tag{PT7.12}$$

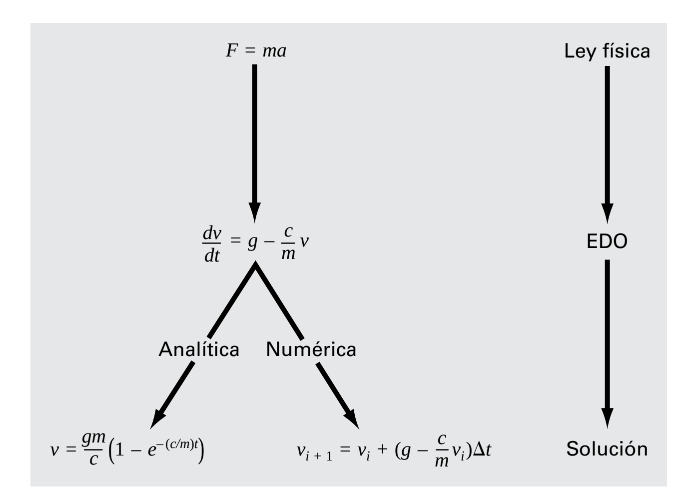


Figura PT7.2
La secuencia de eventos en la aplicación de EDO para resolver problemas de ingeniería. El ejemplo mostrado es la velocidad de un paracaidista en caída.

la cual es un polinomio de cuarto grado (figura PT7.3*a*). Ahora, si derivamos con respecto de *x* a la ecuación (PT7.12), obtenemos una EDO:

$$\frac{dy}{dx} = -2x^3 + 12x^2 - 20x + 8.5 \qquad \text{(PT7.13)}$$

Esta ecuación también describe el comportamiento del polinomio, pero de una manera diferente a la ecuación (PT7.12). Más que representar explícitamente los valores de *y* para cada valor de *x*, la ecuación (PT7.13) da la razón de cambio de *y* con respecto a *x* (es decir, la pendiente) para cada valor de *x*. La figura PT7.3 muestra tanto la función original como la derivada graficadas contra *x*. Observe cómo el valor cero en la derivada corresponde al punto en el cual la función original es plana; es decir, tiene una pendiente cero. Note también que los valores absolutos máximos de la derivada están en los extremos del intervalo donde las pendientes de la función son mayores.

Como se acaba de mostrar, aunque es posible determinar una ecuación diferencial dando la función original, en esencia el objetivo es determinar la función original dada la ecuación diferencial. La función original representa la solución. En el presente caso, esta solución se determina de manera analítica al integrar la ecuación (PT7.13):

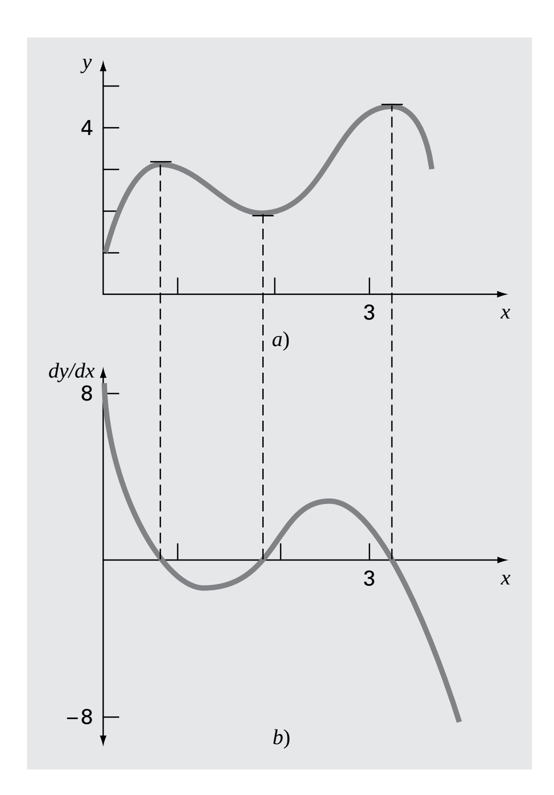


Figura PT7.3
Gráficas de *a*) *y* contra *x*, y *b*) *dy/dx* contra *x* para la función $y = -0.5x^4 + 4x^3 - 10x^2 + 8.5x + 1$.

$$y = \int (-2x^3 + 12x^2 - 20x + 8.5)\, dx$$

Aplicando la regla de integración (recuerde la tabla PT6.2)

$$\int u^n du = \frac{u^{n+1}}{n+1} + C \qquad n \neq -1$$

en cada término de la ecuación, se obtiene la solución

$$y = -0.5x^4 + 4x^3 - 10x^2 + 8.5x + C \qquad \text{(PT7.14)}$$

la cual es idéntica a la función original con una notable excepción. En el proceso de la diferenciación y después en la integración, se pierde el valor constante de 1 en la ecuación original y ganamos el valor *C*. Esta *C* es llamada *constante de integración*. El hecho de que aparezca esta constante arbitraria indica que la solución no es única. Es decir, es solución pero con un número infinito de funciones posibles (correspondiente al número infinito de posibles valores de *C*) que satisfacen la ecuación diferencial. Por ejemplo, la figura PT7.4 muestra seis funciones posibles que satisfacen la ecuación (PT7.14).

Por lo tanto, para especificar la solución por completo, la ecuación diferencial usualmente se encuentra acompañada por *condiciones auxiliares*. Para las EDO de primer orden, se requiere un tipo de condición auxiliar, llamada *valor inicial*, para determinar la constante y obtener una solución única. Por ejemplo, la ecuación (PT7.13) se acompaña por la condición inicial definida por $x = 0$, $y = 1$. Estos valores se sustituyen en la ecuación (PT7.14):

$$1 = -0.5(0)^4 + 4(0)^3 - 10(0)^2 + 8.5(0) + C \qquad \text{(PT7.15)}$$

para determinar $C = 1$. Por consiguiente, la solución única que satisface tanto a la ecuación diferencial como la condición inicial especificada se obtiene al sustituir $C = 1$ en la ecuación (PT7.14):

$$y = -0.5x^4 + 4x^3 - 10x^2 + 8.5x + 1 \tag{PT7.16}$$

De esta forma, hemos "fijado" la ecuación (PT7.14) al forzarla a pasar a través de un punto dado por la condición inicial y, al hacerlo, encontramos una solución única para la EDO y completamos un ciclo con la función original [ecuación (PT7.12)].

Las condiciones iniciales por lo común tienen interpretaciones muy tangibles para las ecuaciones diferenciales surgidas de las condiciones de problemas físicos. Por ejemplo, en el problema del paracaidista en caída, la condición inicial fue tomada del hecho físico de que en el tiempo cero la velocidad vertical es cero. Si el paracaidista hubiese estado en movimiento vertical en el tiempo cero, la solución debería haberse modificado al tomar en cuenta esta velocidad inicial.

Cuando tratamos con una ecuación diferencial de n-ésimo orden, se requiere de n condiciones para obtener una solución única. Si se especifican todas las condiciones en el mismo valor de la variable independiente (por ejemplo, en x o $t = 0$), entonces se conocen como *problemas de valor inicial*. En cambio en los *problemas de valor en la frontera*, la especificación de condiciones ocurre con diferentes valores de la variable independiente. En los capítulos 25 y 26 se analizarán problemas de valor inicial. Los problemas de valor en la frontera se estudiarán en el capítulo 27, junto con los problemas sobre valores propios.

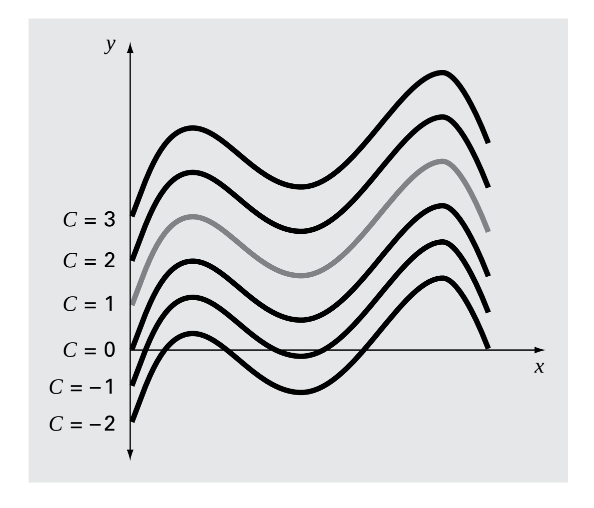


Figura PT7.4
Seis posibles soluciones para la integral de $-2x^3 + 12x^2 - 20x + 8.5$. Cada una corresponde a un valor diferente de la constante de integración C.

PT7.3 ORIENTACIÓN

Antes de proceder con los métodos numéricos para la solución de ecuaciones diferenciales ordinarias, sería de utilidad tener alguna orientación. El siguiente material tiene como propósito ofrecerle una visión general del material que se estudiará en la parte siete. Además, formulamos objetivos para concentrar el análisis sobre el área de estudio.

PT7.3.1 Alcance y presentación preliminar

La figura PT7.5 ofrece un panorama general de la parte siete. Se analizarán en esta parte del libro dos categorías amplias de métodos numéricos, para problemas de valor inicial. Los *métodos de un paso*, los cuales se verán en el capítulo 25, permiten el cálculo y_{i+1}, dada la ecuación diferencial y y_i. Los métodos de pasos múltiples, que se estudiarán en el capítulo 26, requieren valores adicionales de y además de los de y_i.

Todos los métodos, excepto los *métodos de un paso* del *capítulo 25*, pertenecen a lo que se conoce como técnicas de Runge-Kutta. Aunque el capítulo podría haberse organizado alrededor de esta noción teórica, optamos por un procedimiento intuitivo, más gráfico, para presentar los métodos. De esta manera, iniciamos el capítulo con el *método de Euler*, el cual tiene una interpretación gráfica muy directa. Luego se estudian dos versiones mejoradas al método de Euler, que están desarrollados a partir de argumentos visuales (las técnicas de *Heun* y las de *punto medio*). Después de esta introducción, desarrollamos de manera formal el concepto de técnicas de *Runge-Kutta* (o *RK*) y mostramos cómo las técnicas anteriores constituyen el conjunto de métodos RK a partir de los de primero y segundo orden. Continuamos con las formulaciones de los métodos RK de orden superior que se utilizan con mayor frecuencia en la solución de problemas de ingeniería. Además, cubrimos la aplicación de los métodos de un paso en los *sistemas de EDO*. Por último, el capítulo termina con un análisis de los *métodos RK adaptativos* que automáticamente ajustan el tamaño de paso en respuesta al error de truncamiento del cálculo.

El *capítulo 26* inicia con una descripción de las *EDO rígidas*, que se encuentran tanto en forma individual como en los sistemas de EDO, y para su solución ambos tienen componentes lentos y

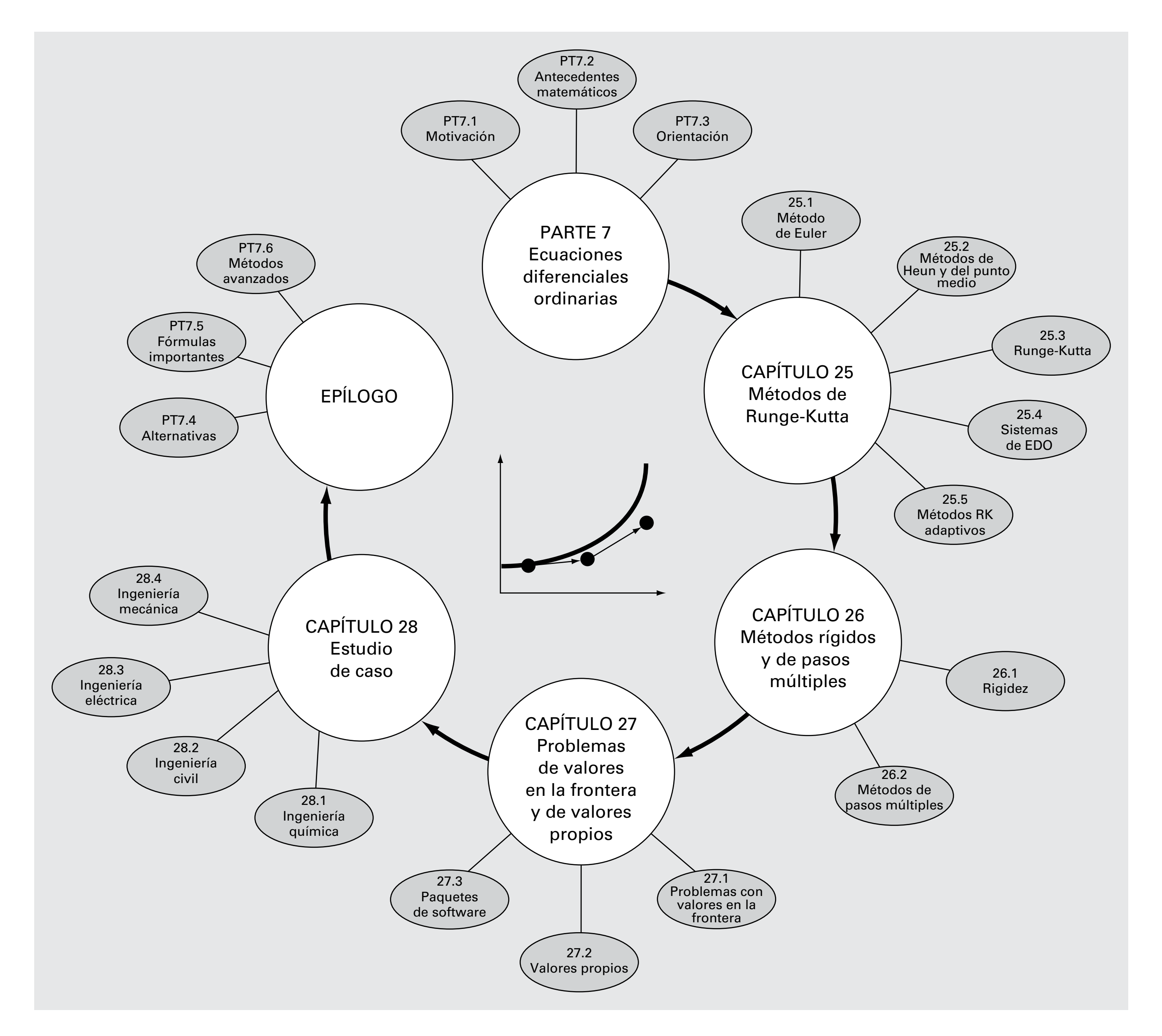


Figura PT7.5
Representación esquemática de la organización de la parte siete: Ecuaciones diferenciales ordinarias.

rápidos. Presentamos la idea de una técnica de *solución implícita* como una de las soluciones más comunes para resolver este problema.

Después, estudiamos los *métodos de pasos múltiples*. Estos algoritmos requieren información de los pasos anteriores para obtener una mayor efectividad en la trayectoria de la solución. También ofrecen una estimación del error de truncamiento que se puede utilizar para implementar un control en el tamaño de paso. En esta sección, seguimos primero un procedimiento intuitivo-visual al usar un método simple (el de *Heun sin autoinicio*), para presentar todas las características esenciales de los procedimientos de pasos múltiples.

En el *capítulo 27* abordamos los problemas de *valores en la frontera* y los problemas de *valores propios* (valores característicos o eigenvalores). Para los primeros, estudiamos tanto los *métodos de disparo* como los de *diferencias finitas*. Para los segundos, analizamos diferentes procedimientos, entre ellos, los *métodos de polinomios* y los *métodos de potencias*. Por último, el capítulo concluye

con una descripción de la aplicación de varios *paquetes de software* y *bibliotecas* para la solución de las EDO y de los valores propios.

El *capítulo 28* se dedica a las aplicaciones en una gran gama de los campos de la ingeniería. Además, se incluye una sección con un breve repaso al final de la parte siete. Este epílogo resume y compara las fórmulas y los conceptos importantes relacionados con las EDO. La comparación incluye un análisis de las ventajas y las desventajas que son relevantes para su implementación en la práctica de la ingeniería. El epílogo también resume fórmulas importantes e incluye referencias sobre temas avanzados.

PT7.3.2 Metas y objetivos

Objetivos de estudio Después de completar la parte siete, usted debe aumentar de manera notoria su capacidad para enfrentar y resolver tanto problemas de ecuaciones diferenciales ordinarias, como de valores propios. Las metas de estudio en general incluyen el manejo de las técnicas, y una capacidad para evaluar la confiabilidad de las respuestas; así como la posibilidad de seleccionar el "mejor" método (o métodos) para cualquier problema en particular. Además de estos objetivos generales, deberán dominar los objetivos de estudio específicos que se muestran en la tabla PT7.2.

Objetivos de cómputo Se le ofrecen los algorítmos para muchos de los métodos en la parte siete. Esta información le permitirá aumentar su biblioteca de software. Por ejemplo, quizá le sea útil desde un punto de vista profesional tener el software que emplea el método de Runge-Kutta de cuarto orden para más de cinco ecuaciones y resolver las EDO con un procedimiento adaptativo de tamaño de paso.

Por último, una de sus más importantes metas deberá ser el dominio de los diversos paquetes de software de propósito general que están disponibles. En particular, deberá convertirse en un entusiasta usuario de esas herramientas para implementar métodos numéricos en la solución de problemas de ingeniería.

Tabla PT7.2 Objetivos específicos de estudio de la parte siete.

1. Comprender las representaciones visuales de los métodos de Euler, de Heun y del punto medio.
2. Conocer la relación del método de Euler con la expansión de la serie de Taylor y la implicación que esto tiene con respecto al error del método.
3. Reconocer la diferencia entre los errores de truncamiento local y global, y cómo se relacionan con la selección de un método numérico para un problema específico.
4. Entender el orden y la dependencia del tamaño de paso respecto de los errores de truncamiento global, para todos los métodos descritos en la parte siete; entender cómo dichos errores tienen que ver con la exactitud de las técnicas.
5. Comprender la base de los métodos predictor-corrector; en particular, percatarse que la eficiencia del corrector es dependiente de la exactitud del predictor.
6. Conocer la forma general de los métodos de Runge-Kutta; entender la deducción del método RK de segundo orden y cómo se relaciona con la expansión de la serie de Taylor; darse cuenta de que hay un número infinito de versiones posibles para los métodos RK de segundo orden y de orden superiores.
7. Saber cómo aplicar cualquiera de los métodos RK a los sistemas de ecuaciones; poder reducir una EDO de n-ésimo orden a un sistema de n-EDO de primer orden.
8. Reconocer el tipo de contexto de un problema donde es importante ajustar el tamaño de paso.
9. Entender cómo se agrega el control del tamaño de paso adaptativo a un método RK de cuarto orden.
10. Saber de qué modo la combinación de los componentes lentos y rápidos actúa en la solución de una ecuación o un sistema de ecuaciones rígidos.
11. Distinguir entre esquemas de solución implícitos y explícitos para las EDO; en particular, reconocer cómo 1) se disminuye la rigidez del problema y 2) se complica la mecánica de solución.
12. Detectar la diferencia entre problemas de valor inicial y de valores en la frontera.
13. Saber la diferencia entre los métodos de pasos múltiples y de un paso; darse cuenta de que todos los métodos de pasos múltiples son predictor-corrector, pero no a la inversa.
14. Comprender la relación entre fórmulas de integración y métodos predictor-corrector.
15. Reconocer la diferencia fundamental entre las fórmulas de integración de Newton-Cotes y la de Adams.
16. Entender la fundamentación de los métodos de polinomios y de potencias para determinar los valores propios; en particular, reconocer sus ventajas y sus limitaciones.
17. Saber cómo la deflación de Hoteller permite que el método de potencias se utilice para calcular los valores propios intermedios.
18. Utilizar los paquetes de software y/o bibliotecas para integrar las EDO y evaluar los valores propios.

CAPÍTULO 25

Métodos de Runge-Kutta

Este capítulo se dedica a la solución de ecuaciones diferenciales ordinarias de la forma

$$\frac{dy}{dx} = f(x, y)$$

En el capítulo 1 se utilizó un método numérico para resolver una ecuación como la anterior, para el cálculo de la velocidad del paracaidista en caída. Recuerde que el método fue de la forma general

Nuevo valor = valor anterior + pendiente × tamaño de paso

o, en términos matemáticos,

$$y_{i+1} = y_i + \phi h \tag{25.1}$$

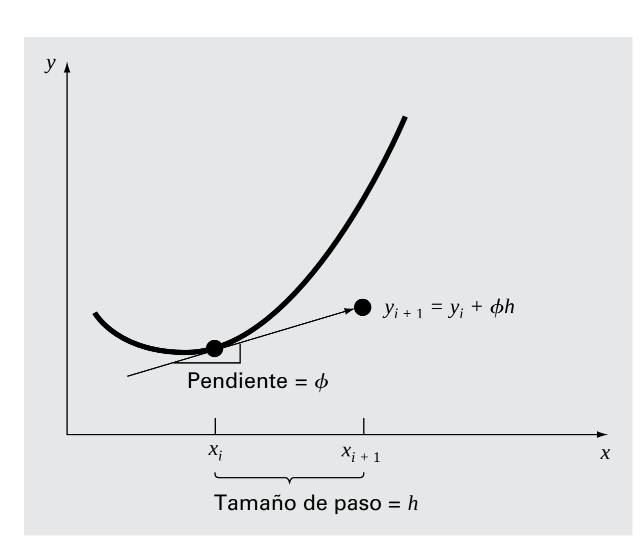


Figura 25.1
Ilustración gráfica del método de un paso.

De acuerdo con esta ecuación, la pendiente estimada ϕ se usa para extrapolar desde un valor anterior y_i a un nuevo valor y_{i+1} en una distancia h (figura 25.1). Esta fórmula se aplica paso a paso para calcular un valor posterior y, por lo tanto, para trazar la trayectoria de la solución.

Todos los métodos de un paso que se expresen de esta forma general, tan sólo van a diferir en la manera en la que se estima la pendiente. Como en el problema del paracaidista en caída, el procedimiento más simple consiste en usar la ecuación diferencial, para estimar la pendiente, en la forma de la primera derivada en x_i. En otras palabras, se toma la pendiente al inicio del intervalo como una aproximación de la pendiente promedio sobre todo el intervalo. Tal procedimiento, llamado *método de Euler*, se analiza en la primera parte de este capítulo. Después se revisan otros métodos de un paso que emplean otras formas de estimar la pendiente que dan como resultado predicciones más exactas. Todas estas técnicas en general se conocen como métodos de *Runge-Kutta*.

25.1 MÉTODO DE EULER

La primera derivada ofrece una estimación directa de la pendiente en x_i (figura 25.2):

$$\phi = f(x_i, y_i)$$

donde $f(x_i, y_i)$ es la ecuación diferencial evaluada en x_i y y_i. La estimación se sustituye en la ecuación (25.1):

$$y_{i+1} = y_i + f(x_i, y_i)h \tag{25.2}$$

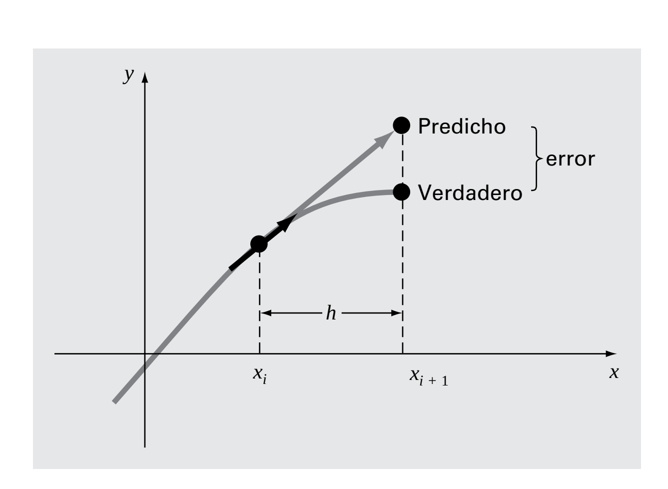


Figura 25.2
Método de Euler.

Esta fórmula se conoce como *método de Euler* (o de *Euler-Cauchy* o de *punto pendiente*). Se predice un nuevo valor de y usando la pendiente (igual a la primera derivada en el valor original de x) para extrapolar linealmente sobre el tamaño de paso h (figura 25.2).

EJEMPLO 25.1 Método de Euler

Planteamiento del problema Con el método de Euler integre numéricamente la ecuación (PT7.13):

$$\frac{dy}{dx} = -2x^3 + 12x^2 - 20x + 8.5$$

desde $x = 0$ hasta $x = 4$ con un tamaño de paso 0.5. La condición inicial en $x = 0$ es $y = 1$. Recuerde que la solución exacta está dada por la ecuación (PT7.16):

$$y = -0.5x^4 + 4x^3 - 10x^2 + 8.5x + 1$$

Solución Se utiliza la ecuación (25.2) para implementar el método de Euler:

$$y(0.5) = y(0) + f(0, 1)0.5$$

donde $y(0) = 1$ y la pendiente estimada en $x = 0$ es:

$$f(0, 1) = -2(0)^3 + 12(0)^2 - 20(0) + 8.5 = 8.5$$

Por lo tanto,

$$y(0.5) = 1.0 + 8.5(0.5) = 5.25$$

La solución verdadera en $x = 0.5$ es:

$$y = -0.5(0.5)^4 + 4(0.5)^3 - 10(0.5)^2 + 8.5(0.5) + 1 = 3.21875$$

Así, el error es:

$$E_t = \text{valor verdadero} - \text{valor aproximado} = 3.21875 - 5.25 = -2.03125$$

o, expresada como error relativo porcentual, $\varepsilon_t = -63.1\%$. En el segundo paso,

$$\begin{aligned} y(1) &= y(0.5) + f(0.5, 5.25)0.5 \\ &= 5.25 + [-2(0.5)^3 + 12(0.5)^2 - 20(0.5) + 8.5]0.5 \\ &= 5.875 \end{aligned}$$

Tabla 25.1 Comparación de los valores verdadero y aproximado de la integral de $y' = -2x^3 + 12x^2 - 20x + 8.5$, con la condición inicial de que $y = 1$ en $x = 0$. Los valores aproximados se calcularon empleando el método de Euler con un tamaño de paso de 0.5. El error local se refiere al error en que se incurre sobre un solo paso. Éste se calcula con una expansión de la serie de Taylor como en el ejemplo 25.2. El error global es la discrepancia total debida a los pasos anteriores y presentes.

			Error relativo porcentual	
x	$y_{verdadero}$	y_{Euler}	Global	Local
0.0	1.00000	1.00000		
0.5	3.21875	5.25000	−63.1	−63.1
1.0	3.00000	5.87500	−95.8	−28.1
1.5	2.21875	5.12500	131.0	−1.4
2.0	2.00000	4.50000	−125.0	20.3
2.5	2.71875	4.75000	−74.7	17.2
3.0	4.00000	5.87500	46.9	3.9
3.5	4.71875	7.12500	−51.0	−11.3
4.0	3.00000	7.00000	−133.3	−53.1

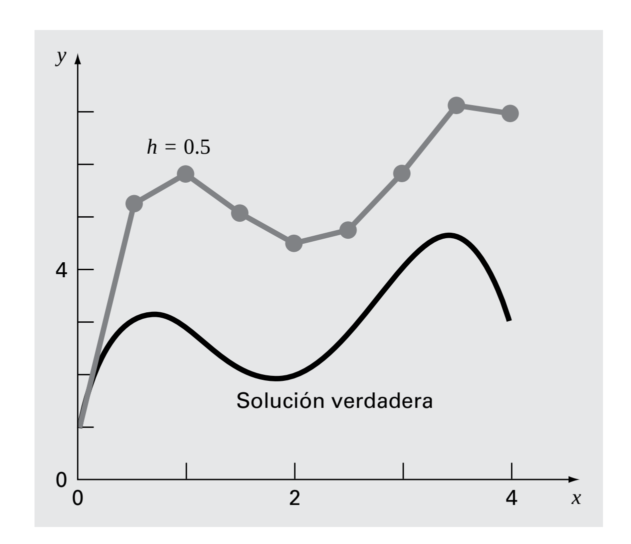


Figura 25.3
Comparación de la solución verdadera con una solución numérica usando el método de Euler, para la integral de $y' = -2x^3 + 12x^2 - 20x + 8.5$ desde $x = 0$ hasta $x = 4$ con un tamaño de paso de 0.5. La condición inicial en $x = 0$ es $y = 1$.

La solución verdadera en $x = 1.0$ es 3.0 y, entonces, el error relativo porcentual es –95.8%. El cálculo se repite y los resultados se dan en la tabla 25.1 y en la figura 25.3. Observe que aunque el cálculo capta la tendencia general de la solución verdadera, el error resulta considerable. Como se explica en la siguiente sección, es posible reducir tal error usando un tamaño de paso menor.

El ejemplo anterior usa un polinomio simple como ecuación diferencial con el objetivo de facilitar el siguiente análisis de error. De esta forma,

$$\frac{dy}{dx} = f(x)$$

En efecto, un caso más general (y más común) implica EDO, donde aparece una función que depende tanto de x como de y,

$$\frac{dy}{dx} = f(x,y)$$

Conforme avancemos en esta parte del texto, nuestros ejemplos comprenderán EDO que dependen de variables independientes y dependientes.

25.1.1 Análisis del error para el método de Euler

La solución numérica de las EDO implica dos tipos de error (recuerde los capítulos 3 y 4):

1. Errores de *truncamiento*, o de discretización, originados por la naturaleza de las técnicas empleadas para aproximar los valores de y.
2. Errores de *redondeo*, causados por el número limitado de cifras significativas que una computadora puede retener.

Los errores de truncamiento se componen de dos partes. La primera es un *error de truncamiento local* que resulta de una aplicación del método considerado, en un solo paso. La segunda es un *error de truncamiento propagado* que resulta de las aproximaciones producidas durante los pasos previos. La suma de los dos es el *error de truncamiento global* o *total*.

Al adquirir cierta comprensión de la magnitud y de las propiedades del error de truncamiento, puede desarrollarse el método de Euler directamente de la expansión de la serie de Taylor. Para ello, observe que la ecuación diferencial que se va a integrar será de la forma general:

$$y' = f(x, y) \tag{25.3}$$

donde $y' = dy/dx$, x y y son las variables independiente y dependiente, respectivamente. Si la solución (es decir, la función que describe el comportamiento de y) tiene derivadas continuas, se representa por una expansión de la serie de Taylor respecto a un valor inicial (x_i, y_i), como sigue [recuerde la ecuación (4.7)]

$$y_{i+1} = y_i + y_i'h + \frac{y_i''}{2!}h^2 + \cdots + \frac{y_i^{(n)}}{n!}h^n + R_n \tag{25.4}$$

donde $h = x_{i+1} - x_i$ y R_n = término remanente, definido como:

$$R_n = \frac{y^{(n+1)}(\xi)}{(n+1)!}h^{n+1} \tag{25.5}$$

donde ξ está en algún lugar en el intervalo de x_i a x_{i+1}. Es posible desarrollar una forma alternativa, sustituyendo la ecuación (25.3) en las ecuaciones (25.4) y (25.5) para obtener:

$$y_{i+1} = y_i + f(x_i, y_i)h + \frac{f'(x_i, y_i)}{2!}h^2 + \cdots + \frac{f^{(n-1)}(x_i, y_i)}{n!}h^n + O(h^{n+1}) \tag{25.6}$$

donde $O(h^{n+1})$ especifica que el error de truncamiento local es proporcional al tamaño de paso elevado a la potencia $(n + 1)$.

Al comparar las ecuaciones (25.2) y (25.6), se advierte que el método de Euler corresponde a la serie de Taylor, hasta el término $f(x_i, y_i)h$ inclusive. Además, la comparación indica que el error de truncamiento se debe a que aproximamos la solución verdadera mediante un número finito de términos de la serie de Taylor. Así, truncamos, o dejamos fuera, una parte de la solución verdadera. Por ejemplo, el error de truncamiento en el método de Euler se atribuye a los términos remanentes en la expansión de la serie de Taylor, que no se incluyeron en la ecuación (25.2). Al restar la ecuación (25.2), que es una aproximación de la (25.6) se llega a

$$E_t = \frac{f'(x_i, y_i)}{2!}h^2 + \cdots + O(h^{n+1}) \tag{25.7}$$

donde E_t = error de truncamiento local verdadero. Para h suficientemente pequeña, los errores en los términos de la ecuación (25.7) normalmente disminuyen, en tanto aumenta el orden (recuerde el ejemplo 4.2 y el análisis que lo acompaña) y el resultado se representa como:

$$E_a = \frac{f'(x_i, y_i)}{2!}h^2 \tag{25.8}$$

o

$$E_a = O(h^2) \tag{25.9}$$

donde E_a = error de truncamiento local aproximado.

EJEMPLO 25.2 Estimación de la serie de Taylor para el error del método de Euler

Planteamiento del problema Con la ecuación (25.7) estime el error en el primer paso del ejemplo 25.1. Úsela también para determinar el error debido a cada uno de los términos de orden superior en la expansión de la serie de Taylor.

Solución Como tenemos un polinomio, se aplica la serie de Taylor para obtener estimaciones exactas del error en el método de Euler. La ecuación (25.7) se escribe como:

$$E_t = \frac{f'(x_i, y_i)}{2!}h^2 + \frac{f''(x_i, y_i)}{3!}h^3 + \frac{f^{(3)}(x_i, y_i)}{4!}h^4 \tag{E25.2.1}$$

donde $f'(x_i, y_i)$ = la primera derivada de la ecuación diferencial (que es la segunda derivada de la solución). En el presente caso,

$$f'(x_i, y_i) = -6x^2 + 24x - 20 \tag{E25.2.2}$$

y $f''(x_i, y_i)$ es la segunda derivada de la EDO

$$f''(x_i, y_i) = -12x + 24 \quad \text{(E25.2.3)}$$

y $f^{(3)}(x_i, y_i)$ es la tercera derivada de la EDO

$$f^{(3)}(x_i, y_i) = -12 \quad \text{(E25.2.4)}$$

Podemos omitir términos adicionales (es decir, la cuarta derivada y las superiores) de la ecuación (E25.2.1), ya que en este caso específico son iguales a cero. Se debe observar que para otras funciones (por ejemplo, funciones trascendentes como senoides o exponenciales) esto no necesariamente es cierto, y los términos de orden superior llegan a tener valores diferentes de cero. Sin embargo, en el caso presente, las ecuaciones (E25.2.1) a la (E25.2.4) definen por completo el error de truncamiento en una sola aplicación del método de Euler.

Por ejemplo, el error debido al truncamiento del término de segundo orden se calcula como sigue:

$$E_{t,2} = \frac{-6(0.0)^2 + 24(0.0) - 20}{2}(0.5)^2 = -2.5 \quad \text{(E25.2.5)}$$

Para el término de tercer orden:

$$E_{t,3} = \frac{-12(0,0) + 24}{6}(0.5)^3 = 0.5$$

y para el término de cuarto orden:

$$E_{t,4} = \frac{-12}{24}(0.5)^4 = -0.03125$$

Se suman los tres resultados para obtener el error total de truncamiento:

$$E_t = E_{t,2} + E_{t,3} + E_{t,4} = -2.5 + 0.5 - 0.03125 = -2.03125$$

que es exactamente el error en que se incurrió en el paso inicial del ejemplo 25.1. Observe cómo $E_{t,2} > E_{t,3} > E_{t,4}$, lo cual justifica la aproximación representada por la ecuación (25.8).

Como se ilustra en el ejemplo 25.2, la serie de Taylor ofrece un medio de cuantificar el error en el método de Euler. Aunque existen limitaciones asociadas con su empleo para tal propósito:

1. La serie de Taylor permite sólo una estimación del error de truncamiento local; es decir, el error generado durante un solo paso del método. No ofrece una medida del error propagado, por lo tanto, ni del error de truncamiento global. En la tabla 25.1 se incluyen los errores de truncamiento local y global para el ejemplo 25.1. El error local se calculó en cada paso con la ecuación (25.2), pero usando el valor verdadero de y_i (la segunda columna de la tabla) para calcular y_{i+1} y no el valor aproximado (la tercera columna), como se hizo con el método de Euler. Como se esperaba, el error de truncamiento local absoluto promedio (25%) es menor que el error global promedio (90%). La única razón por la que es posible realizar estos cálculos de error exactos es que conocemos *a priori* el valor verdadero. Obviamente éste no será el caso en un problema real. En consecuencia, como lo analizaremos después, usted a menudo debe aplicar técnicas como el método de Euler, usando varios tamaños de paso, para obtener una estimación indirecta de los errores.
2. Como se mencionó anteriormente, en problemas reales por lo común se tienen funciones más complicadas que simples polinomios. En consecuencia, las derivadas que se necesitan para obtener la expansión de la serie de Taylor no siempre serán fáciles de calcular.

Aunque estas limitaciones impiden el análisis exacto del error en la mayoría de los problemas prácticos, la serie de Taylor brinda una valiosa ayuda en la comprensión en el comportamiento del método de Euler. De acuerdo con la ecuación (25.9), se advierte que el error local es proporcional al cuadrado del tamaño de paso y a la primera derivada de la ecuación diferencial. También se puede demostrar que el error de truncamiento global es $O(h)$; es decir, es proporcional al tamaño de paso (Carnahan y colaboradores, 1969). Estas observaciones permiten establecer las siguientes conclusiones útiles:

1. Se puede reducir el error disminuyendo el tamaño del paso.
2. El método dará como resultado predicciones sin error si la función que se analiza (es decir, la solución de la ecuación diferencial) es lineal, debido a que en una línea recta la segunda derivada es cero.

Esta última conclusión tiene un sentido intuitivo, puesto que el método de Euler usa segmentos de línea recta para aproximar la solución. De ahí que al método de Euler se le conozca como un *método de primer orden*.

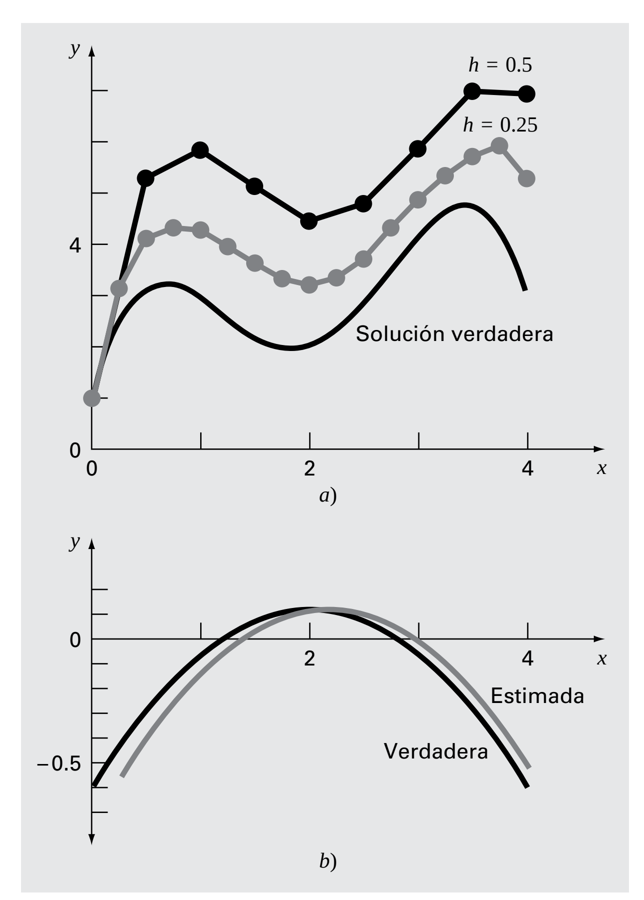
También deberá observarse que este patrón general rige a los métodos de orden superior de un paso, que se describen en las siguientes páginas. Es decir, un método de n-ésimo orden dará resultados perfectos si la solución de la EDO es un polinomio de n-ésimo grado. Además, el error de truncamiento local será $O(h^{n+1})$; y el error global, $O(h^n)$.

EJEMPLO 25.3 Efecto de un tamaño de paso reducido en el método de Euler

Planteamiento del problema Repita el cálculo del ejemplo 25.1, pero ahora use un tamaño de paso igual a 0.25.

Solución Los cálculos se repiten, y los resultados se recopilan en la figura 25.4*a*). Al reducir el tamaño de paso a la mitad, el valor absoluto del error global promedio disminuye al 40%, y el valor absoluto del error local al 6.4%. Esto se compara con los errores global y local del ejemplo 25.1, 90 y 24.8%, respectivamente. Así, como se esperaba, el error local disminuye a un cuarto y el error global a la mitad.

Observe también cómo el error local cambia de signo en valores intermedios a lo largo del intervalo, lo cual se debe principalmente a que la primera derivada de la ecuación diferencial es una parábola que cambia de signo [recuerde la ecuación (E25.2.2) y examine la figura 25.4*b*]. Debido a que el error local es proporcional a esta función, el efecto total de la oscilación en el signo es evitar un crecimiento continuo del error global conforme se ejecuta el cálculo. Así, desde $x = 0$ hasta $x = 1.25$, todos los errores locales son negativos y, en consecuencia, el error global aumenta en este intervalo. En la sección intermedia del intervalo, los errores locales positivos comienzan a reducir el error global. Cerca del extremo, se invierte el proceso y, de nuevo, aumenta el error global. Si el error local continuamente cambia de signo sobre el intervalo de cálculo, normalmente el efecto total es reducir el error global. No obstante, si los errores locales son del mismo signo, entonces la solución numérica puede divergir cada vez más de la solución verdadera, en tanto se ejecuta el cálculo. Se dice que tales resultados son *inestables*.



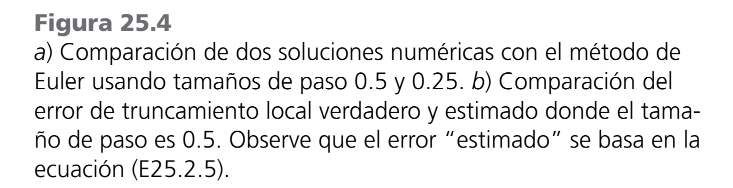


Figura 25.4
a) Comparación de dos soluciones numéricas con el método de Euler usando tamaños de paso 0.5 y 0.25. *b*) Comparación del error de truncamiento local verdadero y estimado donde el tamaño de paso es 0.5. Observe que el error "estimado" se basa en la ecuación (E25.2.5).

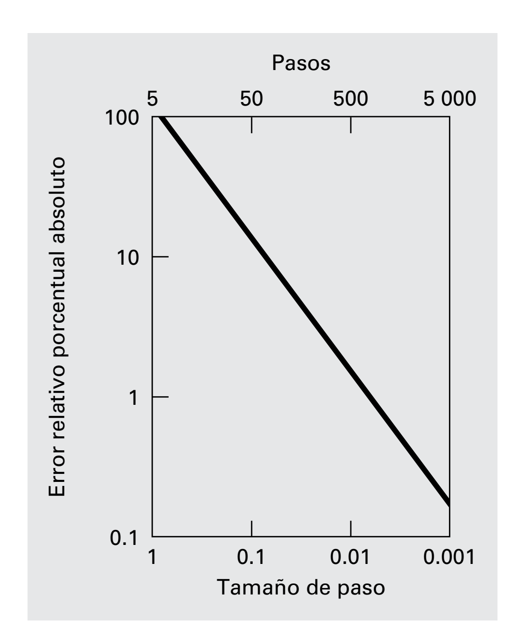


Figura 25.5
Efecto del tamaño de paso sobre el error de truncamiento global en el método de Euler para la integral de $y' = -2x^3 + 12x^2 - 20x + 8.5$. La gráfica muestra el error global relativo porcentual absoluto en $x = 5$ en función del tamaño de paso.

El efecto de algunas reducciones del tamaño de paso sobre el error de truncamiento global del método de Euler se ilustra en la figura 25.5, esta gráfica muestra el error relativo porcentual absoluto en $x = 5$ en función del tamaño de paso para el problema que se estudió en los ejemplos 25.1 a 25.3. Observe que aun cuando h se reduce a 0.001, el error todavía es mayor de 0.1%. Ya que este tamaño de paso corresponde a 5 000 pasos para ir desde $x = 0$ hasta $x = 5$, la gráfica sugiere que una técnica de primer orden, como el método de Euler, requiere de muchos cálculos para obtener niveles de error aceptables. Más adelante en este capítulo, se presentarán técnicas de orden superior que dan mucha mayor exactitud con el mismo trabajo de cálculo. Sin embargo, deberá observarse que, a pesar de su ineficiencia, la simplicidad del método de Euler lo hace una opción extremadamente atractiva para muchos problemas de ingeniería. Puesto que es muy fácil de programar, en particular la técnica es útil para llevar a cabo análisis rápidos. En la próxima sección se desarrolla un algoritmo computacional para el método de Euler.

25.1.2 Algoritmo para el método de Euler

Los algoritmos para las técnicas de un paso como el método de Euler son muy simples de programar. Como se especificó al inicio de este capítulo, todos los métodos de un paso tienen la forma general

$$\text{Nuevo valor} = \text{valor anterior} + \text{pendiente} \times \text{tamaño de paso} \tag{25.10}$$

En lo único que difieren los métodos es en el cálculo de la pendiente.

Suponga que usted quiere realizar el cálculo simple expuesto en la tabla 25.1. Es decir, a usted le gustaría utilizar el método de Euler para integrar $y' = -2x^3 + 12x^2 - 20x + 8.5$, con la condición inicial de que $y = 1$ en $x = 0$. Usted quiere integrarla hasta $x = 4$ usando un tamaño de paso de 0.5, y desplegar todos los resultados. Un pseudocódigo simple para realizar esto será como el de la figura 25.6.

Aunque este programa "hará el trabajo" de duplicar los resultados de la tabla 25.1 no está muy bien diseñado. Primero, y ante todo, no es muy modular. Aunque esto no es muy importante para un programa así de pequeño, podría resultar crítico si deseamos modificar y mejorar el algoritmo.

Además, existen algunos detalles relacionados con la forma en que se establecen las iteraciones. Por ejemplo, suponga que el tamaño de paso se volviese muy pequeño para obtener mayor exactitud. En tales casos, debido a que se despliega cada valor calculado, la cantidad de valores de salida podría ser muy grande. Asimismo, el algoritmo supone que el intervalo de cálculo es divisible entre el tamaño de paso. Por último, la acumulación de x en la línea $x = x + dx$ puede estar sujeta a la cuantificación de errores analizada en la sección 3.4.1. Por ejemplo, si dx se cambiara a 0.01 y se usara la representación estándar IEEE de punto flotante (cerca de siete cifras significativas), el resultado al final del cálculo sería 3.999997 en lugar de 4. Para $dx = 0.001$, ¡sería 3.999892!

En la figura 25.7 se muestra un algoritmo mucho más modular que evita esas dificultades. El algoritmo no despliega todos los valores calculados. En lugar de eso, el usuario especifica un intervalo de salida, *xout*, que indica el intervalo en el cual los resultados calculados se guardan en arreglos, xp_m y yp_m. Dichos valores se guardan en arreglos, de tal modo que se puedan desplegar de diferentes formas una vez que termine el cálculo (por ejemplo, impresos graficados, escritos en un archivo).

```
'intervalo de integración
xi = 0
xf = 4
'variables iniciales
x = xi
y = 1
'establece el tamaño de paso y determina el
'número de pasos de cálculo
dx = 0.5
nc = (xf - xi)/dx
'condiciones de salida inicial
PRINT x, y
'ciclo para implementar el método de Euler
'y despliegue de resultados
DOFOR i = 1, nc
  dydx = -2x^3 + 12x^2 - 20x + 8.5
  y = y + dydx · dx
  x = x + dx
  PRINT x, y
END DO
```

Figura 25.6
Pseudocódigo para una primera versión del método de Euler.

a) **Programa principal o "manejador"**

```
Asigna valores para
y = valor inicial variable dependiente
xi = valor inicial variable independiente
xf = valor final variable independiente
dx = cálculo del tamaño de paso
xout = intervalo de salida

x = xi
m = 0
xp_m = x
yp_m = y
DO
  xend = x + xout
  IF (xend > xf) THEN xend = xf
  h = dx
  CALL Integrator (x, y, h, xend)
  m = m + 1
  xp_m = x
  yp_m = y
  IF (x ≥ xf) EXIT
END DO
DISPLAY RESULTS
END
```

b) **Rutina para tomar un paso de salida**

```
SUB Integrator (x, y, h, xend)
  DO
    IF (xend - x < h) THEN h = xend - x
    CALL Euler (x, y, h, ynew)
    y = ynew
    IF (x ≥ xend) EXIT
  END DO
END SUB
```

c) **Método de Euler para una EDO sola**

```
SUB Euler (x, y, h, ynew)
  CALL Derivs(x, y, dydx)
  ynew = y + dydx * h
  x = x + h
END SUB
```

d) **Rutina para determinar la derivada**

```
SUB Derivs (x, y, dydx)
  dydx = ...
END SUB
```

Figura 25.7
Pseudocódigo para una versión modular "mejorada" del método de Euler.

El programa principal toma grandes pasos de salida y llama a una rutina denominada Integrator que hace los pasos de cálculo más pequeños. Observe que los ciclos o *loops* que controlan tanto los pasos grandes como los pequeños terminan basándose en condiciones lógicas. Así, los intervalos no tienen que ser divisibles entre los tamaños de paso.

La rutina Integrator llama después a la rutina Euler que realiza un solo paso con el método de Euler. La rutina Euler llama a una rutina Derivate que calcula el valor de la derivada.

Aunque parecería que tal forma modular es demasiada para el presente caso, facilitará en gran medida la modificación del programa posteriormente. Por ejemplo, aunque el programa de la figura 25.7 está diseñado específicamente para implementar el método de Euler, el módulo de Euler es la única parte que es específica para el método. Así, todo lo que se requiere para aplicar este algoritmo a otros métodos de un paso es modificar esta rutina.

EJEMPLO 25.4 Solución de una EDO con la computadora

Planteamiento del problema Es factible desarrollar un programa computacional a partir del pseudocódigo de la figura 25.7. Este software también sirve para resolver otro problema relacionado con la caída del paracaidista. Usted recordará de la parte uno que nuestro modelo matemático para la velocidad se basaba en la segunda ley de Newton, escrita en la forma:

$$\frac{dv}{dt} = g - \frac{c}{m}v \tag{E25.4.1}$$

Esta ecuación diferencial se resolvió tanto de manera analítica (ejemplo 1.1) como numérica usando el método de Euler (ejemplo 1.2). Las soluciones fueron para el caso donde $g = 9.81$, $c = 12.5$, $m = 68.1$ y $v = 0$ en $t = 0$.

El objetivo del presente ejemplo es repetir esos cálculos numéricos empleando un modelo más complicado para la velocidad con base en una descripción matemática más completa de la fuerza de arrastre causada por la resistencia del viento. Este modelo se obtiene como:

$$\frac{dv}{dt} = g - \frac{c}{m}\left[v + a\left(\frac{v}{v_{\text{máx}}}\right)^b\right] \tag{E25.4.2}$$

donde g, m y c son las mismas que en la ecuación (E25.4.1), a, b y $v_{\text{máx}}$ son constantes empíricas, las cuales, en este caso, son iguales a 8.3, 2.2 y 46, respectivamente. Observe que con este modelo se puede ajustar con exactitud las mediciones empíricas de fuerzas de arrastre contra la velocidad, mejor que el modelo lineal simple del ejemplo 1.1. Sin embargo, esta flexibilidad mayor se gana a expensas de evaluar tres coeficientes en lugar de uno. Además, el modelo matemático resultante es más difícil de resolver en forma analítica. En este caso, el método de Euler ofrece una alternativa conveniente para obtener una solución numérica aproximada.

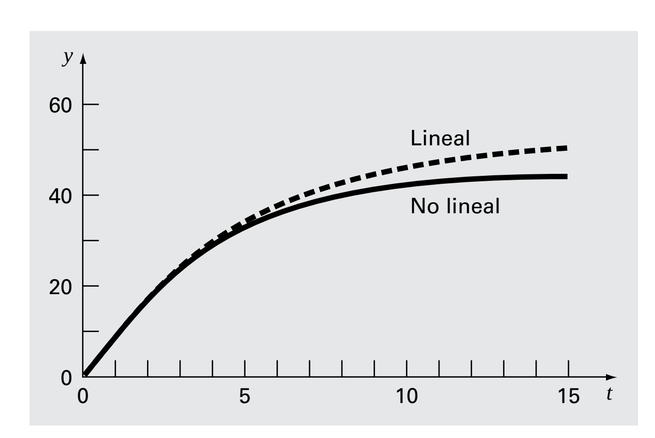


Figura 25.8
Resultados gráficos para la solución de la EDO no lineal [ecuación (E25.4.2)]. Observe que la gráfica también muestra la solución para el modelo lineal [ecuación (E25.4.1)] con propósitos de comparación.

Solución Los resultados para ambos modelos, lineal y no lineal, se muestran en la figura 25.8 con un tamaño de paso de integración de 0.1 s. La gráfica de la figura presenta también una coincidencia de la solución del modelo lineal con propósitos de comparación.

Los resultados de las dos simulaciones indican cómo el aumento en la complejidad de la formulación para la fuerza de arrastre afecta la velocidad del paracaidista. En este caso, la velocidad terminal disminuye debido a la resistencia causada por los términos de orden superior agregados en la ecuación (E25.4.2).

En forma similar es posible utilizar modelos alternativos. La combinación de una solución generada con la computadora vuelve esto una tarea fácil y eficiente. Tales beneficios le permitirán dedicar más tiempo a considerar alternativas creativas y aspectos holísticos del problema, en lugar de los tediosos cálculos a mano.

25.1.3 Métodos para la serie de Taylor de orden superior

Una manera de reducir el error con el método de Euler sería incluir términos de orden superior en la expansión de la serie de Taylor para la solución. Por ejemplo, al incluir el término de segundo orden en la ecuación (25.6) resulta:

$$y_{i+1} = y_i + f(x_i, y_i)h + \frac{f'(x_i, y_i)}{2!}h^2 \tag{25.11}$$

con un error de truncamiento aproximado de:

$$E_a = \frac{f''(x_i, y_i)}{6}h^3$$

Aunque la incorporación de términos de orden superior es simple para implementarse en los polinomios, su inclusión no es tan trivial cuando la EDO es más complicada. En particular, las EDO que están en función tanto de la variable dependiente como de la independiente requieren de la derivación usando la regla de la cadena. Por ejemplo, la primera derivada de $f(x, y)$ es:

$$f'(x_i, y_i) = \frac{\partial f(x,y)}{\partial x} + \frac{\partial f(x,y)}{\partial y}\frac{dy}{dx}$$

La segunda derivada es:

$$f''(x_i, y_i) = \frac{\partial[\partial f/\partial x + (\partial f/\partial y)(dy/dx)]}{\partial x} + \frac{\partial[\partial f/\partial x + (\partial f/\partial y)(dy/dx)]}{\partial y}\frac{dy}{dx}$$

Las derivadas de orden superior se van haciendo cada vez más complicadas.

En consecuencia, como se describe en las próximas secciones, se han desarrollado métodos alternativos de un paso. Estos esquemas son comparables en eficiencia con los procedimientos de la serie de Taylor de orden superior, aunque requieren sólo del cálculo de las primeras derivadas.

25.2 MEJORAS DEL MÉTODO DE EULER

Un motivo fundamental de error en el método de Euler es suponer que la derivada al inicio del intervalo es la misma durante todo el intervalo. Hay dos modificaciones simples para evitar esta consideración. Como se demostrará en la sección 25.3, de hecho, ambas modificaciones pertenecen a una clase superior de técnicas de solución llamadas métodos de Runge-Kutta. Debido a que tienen una interpretación gráfica muy directa, los presentaremos antes de una deducción formal como los métodos de Runge-Kutta.

25.2.1 Método de Heun

Un método para mejorar la estimación de la pendiente emplea la determinación de dos derivadas en el intervalo (una en el punto inicial y otra en el final). Las dos derivadas se promedian después con la finalidad de obtener una mejor estimación de la pendiente en todo el intervalo. Este procedimiento, conocido como *método de Heun*, se presenta en forma gráfica en la figura 25.9.

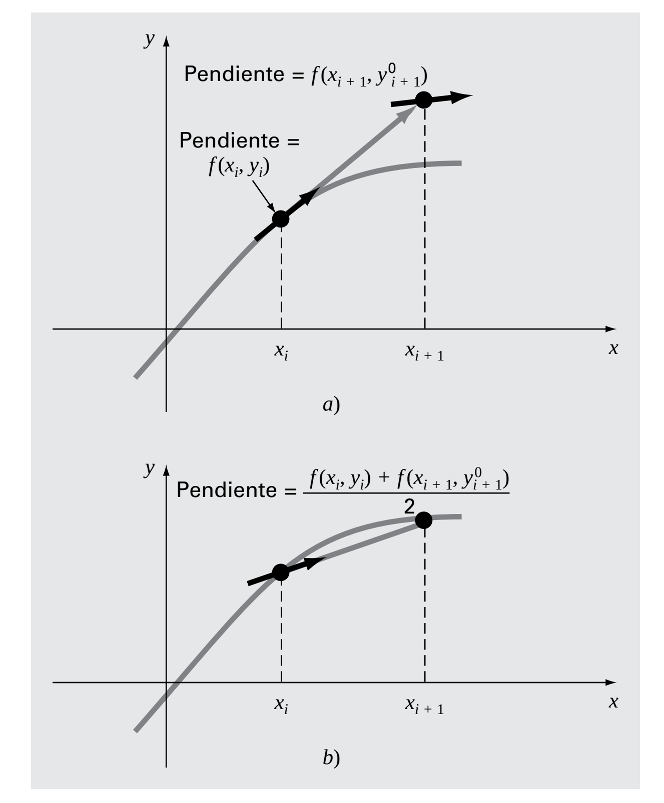


Figura 25.9
Representación gráfica del método de Heun. *a*) Predictor y *b*) corrector.

Recuerde que en el método de Euler, la pendiente al inicio de un intervalo

$$y'_i = f(x_i, y_i) \tag{25.12}$$

se utiliza para extrapolar linealmente a y_{i+1}:

$$y^0_{i+1} = y_i + f(x_i, y_i)h \tag{25.13}$$

En el método estándar de Euler se debería parar aquí. Sin embargo, en el método de Heun la y^0_{i+1} calculada en la ecuación (25.13) no es la respuesta final, sino una predicción intermedia. Por consiguiente, la distinguimos con un superíndice 0. La ecuación (25.13) se llama *ecuación predictora* o simplemente predictor. Da una estimación de y_{i+1} que permite el cálculo de una estimación de la pendiente al final del intervalo:

$$y'_{i+1} = f(x_{i+1}, y^0_{i+1}) \tag{25.14}$$

Así, se combinan las dos pendientes [ecuaciones (25.12) y (25.14)] para obtener una pendiente promedio en el intervalo:

$$\bar{y}' = \frac{y'_i + y'_{i+1}}{2} = \frac{f(x_i, y_i) + f(x_{i+1}, y^0_{i+1})}{2}$$

Esta pendiente promedio se utiliza después para extrapolar linealmente desde y_i hasta y_{i+1} con el método de Euler:

$$y_{i+1} = y_i + \frac{f(x_i, y_i) + f(x_{i+1}, y^0_{i+1})}{2} h$$

que se conoce como *ecuación correctora* o simplemente *corrector.*

El método de Heun es un *procedimiento predictor-corrector.* Todos los métodos de pasos múltiples que se analizarán más adelante en el capítulo 26 son de este tipo. El método de Heun es el único método predictor-corrector de un solo paso que se describe en este libro. Como se desarrolló antes, se expresa en forma concisa como:

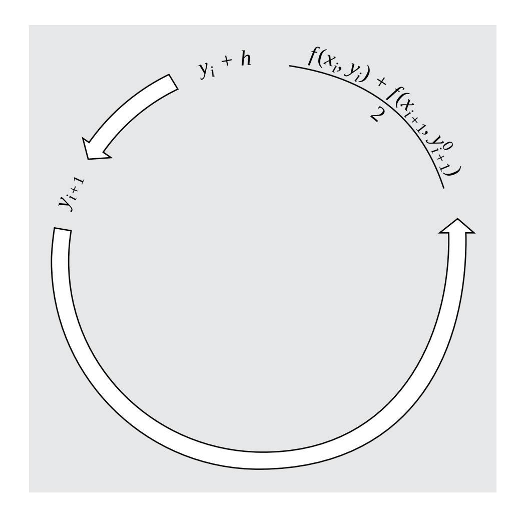


Figura 25.10
Representación gráfica de la forma iterativa del método corrector de Heun para obtener una mejor estimación.

Predictor (figura 25.9*a*) $y_{i+1}^0 = y_i + f(x_i, y_i)h$ (25.15)

Corrector (figura 25.9*b*) $y_{i+1} = y_i + \dfrac{f(x_i, y_i) + f(x_{i+1}, y_{i+1}^0)}{2} h$ (25.16)

Observe que debido a que en la ecuación (25.16) aparece y_{i+1} a ambos lados del signo igual, entonces se puede aplicar en una forma iterativa. Es decir, una estimación anterior se utilizará de manera repetida para proporcionar una estimación mejorada de y_{i+1}. El proceso se ilustra en la figura 25.10. Deberá entenderse que este proceso iterativo no necesariamente converge a la respuesta verdadera, sino que lo hará a una estimación con un error de truncamiento finito, como se mostrará en el siguiente ejemplo.

Como en los métodos iterativos similares analizados en secciones anteriores de este libro, un criterio de terminación para la convergencia del corrector está dado por [recuerde la ecuación (3.5)]

$$|\varepsilon_a| = \left| \frac{y_{i+1}^j - y_{i+1}^{j-1}}{y_{i+1}^j} \right| 100\% \tag{25.17}$$

donde y_{i+1}^{j-1} y y_{i+1}^j resultan de las iteraciones anterior y actual del corrector, respectivamente.

EJEMPLO 25.5 Método de Heun

Planteamiento del problema Con el método de Heun integre $y' = 4e^{0.8x} - 0.5y$ desde $x = 0$ hasta $x = 4$, con un tamaño de paso igual a 1. La condición inicial es en $x = 0$, $y = 2$.

Solución Antes de resolver el problema numéricamente, se utiliza el cálculo para determinar la siguiente solución analítica:

$$y = \frac{4}{1.3}(e^{0.8x} - e^{-0.5x}) + 2e^{-0.5x} \tag{E25.5.1}$$

Tabla 25.2 Comparación de los valores verdadero y aproximado para la integral de $y' = 4e^{0.8x} - 0.5y$, con la condición inicial $y = 2$ en $x = 0$. Los valores aproximados se calcularon utilizando el método de Heun con un tamaño de paso igual a 1. Se muestran dos aplicaciones que corresponden a números diferentes de iteraciones del corrector, junto con el error relativo porcentual absoluto.

		Iteraciones del método de Heun			
		1		15	
x	$y_{verdadero}$	y_{Heun}	$\|\varepsilon_t\|$(%)	y_{Heun}	$\|\varepsilon_t\|$(%)
0	2.0000000	2.0000000	0.00	2.0000000	0.00
1	6.1946314	6.7010819	8.18	6.3608655	2.68
2	14.8439219	16.3197819	9.94	15.3022367	3.09
3	33.6771718	37.1992489	10.46	34.7432761	3.17
4	75.3389626	83.3377674	10.62	77.7350962	3.18

Esta fórmula sirve para generar los valores de la solución verdadera en la tabla 25.2.

Primero, se calcula la pendiente en (x_0, y_0) como

$$y_0' = 4e^0 - 0.5(2) = 3$$

Este resultado difiere mucho de la pendiente promedio real en el intervalo de 0 a 1.0, que es igual a 4.1946, como se calculó de la ecuación diferencial usando la ecuación (PT6.4).

La solución numérica se obtiene al usar el predictor [ecuación (25.15)] para llegar a un estimado de y en 1.0:

$$y_1^0 = 2 + 3(1) = 5$$

Observe que éste es el resultado obtenido con el método estándar de Euler. El valor verdadero en la tabla 25.2 muestra que corresponde a un error relativo porcentual del 19.3%.

Ahora, para mejorar el estimado de y_{i+1}, se emplea el valor y_1^0 para predecir la pendiente al final del intervalo:

$$y_1' = f(x_1, y_1^0) = 4e^{0.8(1)} - 0.5(5) = 6.402164$$

que se combina con la pendiente inicial para obtener una pendiente promedio en el intervalo desde $x = 0$ hasta 1:

$$y' = \frac{3 + 6.402164}{2} = 4.701082$$

que está más cerca a la pendiente promedio verdadera, 4.1946. Dicho resultado se sustituye en el corrector [ecuación (25.16)] para obtener la predicción en $x = 1$:

$$y_1 = 2 + 4.701082(1) = 6.701082$$

que representa un error relativo porcentual de –8.18%. Así, el método de Heun sin iteración del corrector reduce el valor absoluto del error en un factor de 2.4 en comparación con el método de Euler.

Ahora dicho estimado se utiliza para mejorar o corregir la predicción de y_1 sustituyendo el nuevo resultado en el lado derecho de la ecuación (25.16):

$$y_1 = 2 + \frac{[3 + 4e^{0.8(1)} - 0.5(6.701082)]}{2}1 = 6.275811$$

que representa un error relativo porcentual absoluto del 1.31%. El resultado, a su vez, se sustituye en la ecuación (25.16) para corregir aún más:

$$y_1 = 2 + \frac{[3 + 4e^{0.8(1)} - 0.5(6.275811)]}{2}1 = 6.382129$$

que representa un $|\varepsilon_t|$ de 3.03%. Observe cómo los errores algunas veces crecen conforme se llevan a cabo las iteraciones. Tales incrementos pueden ocurrir especialmente con grandes tamaños de paso, y nos previenen de llegar a una conclusión general errónea de que siempre una iteración más mejorará el resultado. No obstante, con tamaños de paso lo suficientemente pequeños, las iteraciones, a la larga, deberán converger a un solo valor. En nuestro caso, 6.360865, que representa un error relativo de 2.68%, que se obtiene después de 15 iteraciones. La tabla 25.2 presenta los resultados del resto de los cálculos usando el método con 1 y 15 iteraciones por paso.

En el ejemplo anterior, la derivada es una función tanto de la variable dependiente y como de la variable independiente x. En situaciones como los polinomios, donde la EDO es únicamente función de la variable independiente, no se requiere el paso predictor [ecuación (25.16)] y el corrector se aplica sólo una vez en cada iteración. En tales casos, la técnica se expresa en forma concisa como sigue:

$$y_{i+1} = y_i + \frac{f(x_i) + f(x_{i+1})}{2}h \tag{25.18}$$

Observe la similitud entre el lado derecho de la ecuación (25.18) y la regla del trapecio [ecuación (21.3)]. La relación entre los dos métodos se puede demostrar formalmente empezando con la ecuación diferencial ordinaria:

$$\frac{dy}{dx} = f(x)$$

De esta ecuación se obtiene y por integración:

$$\int_{y_i}^{y_{i+1}} dy = \int_{x_i}^{x_{i+1}} f(x)\,dx \tag{25.19}$$

cuyo resultado es:

$$y_{i+1} - y_i = \int_{x_i}^{x_{i+1}} f(x)\,dx \tag{25.20}$$

o

$$y_{i+1} = y_i + \int_{x_i}^{x_{i+1}} f(x)\,dx \tag{25.21}$$

Ahora, del capítulo 21 recuerde que la regla del trapecio [ecuación (21.3)] se define como:

$$\int_{x_i}^{x_{i+1}} f(x)\,dx \cong \frac{f(x_i) + f(x_{i+1})}{2} h \tag{25.22}$$

donde $h = x_{i+1} - x_i$. Sustituyendo la ecuación (25.22) en la (25.21) se tiene:

$$y_{i+1} = y_i + \frac{f(x_i) + f(x_{i+1})}{2} h \tag{25.23}$$

que es equivalente a la ecuación (25.18).

Como la ecuación (25.23) es una expresión directa de la regla del trapecio, el error de truncamiento local está dado por [recuerde la ecuación (21.6)]:

$$E_t = -\frac{f''(\xi)}{12} h^3 \tag{25.24}$$

donde ξ está entre x_i y x_{i+1}. Así, el método es de segundo orden porque la segunda derivada de la EDO es cero mientras que la solución verdadera es una cuadrática. Además, los errores local y global son $O(h^3)$ y $O(h^2)$, respectivamente. Entonces, al disminuir el tamaño de paso se disminuye el error más rápido que en el método de Euler. La figura 25.11, que muestra el resultado usando el método de Heun para resolver el polinomio del ejemplo 25.1, demuestra dicho comportamiento.

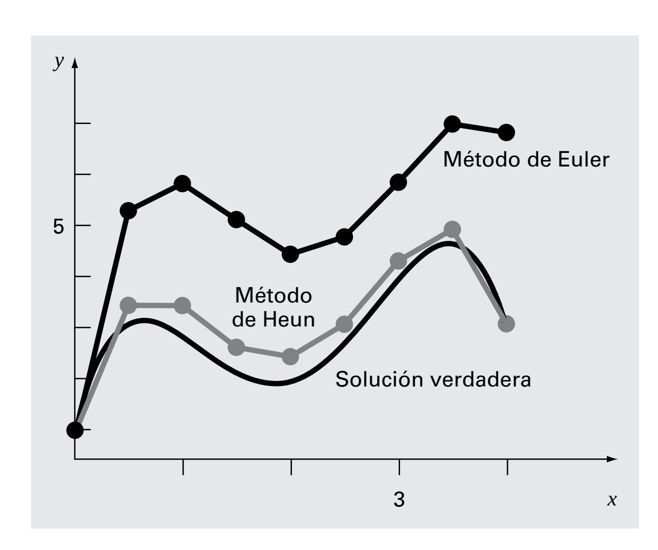


Figura 25.11
Comparación de la solución verdadera con la solución numérica usando los métodos de Euler y de Heun para la integral de $y' = -2x^3 + 12x^2 - 20x + 8.5$.

25.2.2 Método del punto medio (o del polígono mejorado)

La figura 25.12 ilustra otra modificación simple del método de Euler. Conocida como *método del punto medio* (o *del polígono mejorado* o *el modificado de Euler*), esta técnica usa el método de Euler para predecir un valor de y en el punto medio del intervalo (figura 25.12*a*):

$$y_{i+1/2} = y_i + f(x_i, y_i)\frac{h}{2} \tag{25.25}$$

Después, este valor predicho se utiliza para calcular una pendiente en el punto medio:

$$y'_{i+1/2} = f(x_{i+1/2}, y_{i+1/2}) \qquad (25.26)$$

que se supone representa una aproximación válida de la pendiente promedio en todo el intervalo. Dicha pendiente se usa después para extrapolar linealmente desde x_i hasta x_{i+1} (figura 25.12*b*):

$$y_{i+1} = y_i + f(x_{i+1/2}, y_{i+1/2})h \qquad (25.27)$$

Observe que como y_{i+1} no está en los dos lados, el corrector [ecuación (25.27)] no puede aplicarse de manera iterativa para mejorar la solución.

Como en la sección anterior, este procedimiento también se relaciona con las fórmulas de integración de Newton-Cotes. Recuerde de la tabla 21.4 que la fórmula más simple de integración abierta de Newton-Cotes, la cual se denomina el método del punto medio, se puede representar como:

$$\int_a^b f(x)\,dx \cong (b-a)f(x_1)$$

donde x_1, es el punto medio del intervalo (a, b). Usando la nomenclatura del caso actual, lo anterior se expresa como:

$$\int_{x_i}^{x_{i+1}} f(x)\,dx \cong hf(x_{i+1/2})$$

La sustitución de esta fórmula en la ecuación (25.21) dará la ecuación (25.27). De esta manera, así como al método de Heun se le puede llamar la regla del trapecio, el *método del punto medio* obtiene su nombre de la fórmula de integración correspondiente sobre la cual se basa.

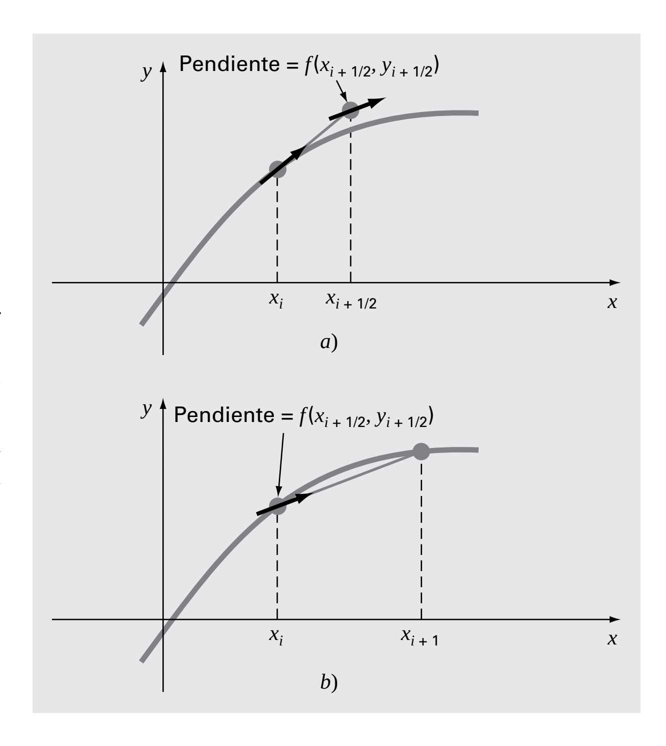


Figura 25.12
Representación gráfica del método del punto medio. *a*) Ecuación (25.25) y *b*) ecuación (25.27).

El método del punto medio es mejor que el método de Euler debido a que utiliza una estimación de la pendiente en el punto medio del intervalo de predicción. Recuerde que, como vimos en la diferenciación numérica de la sección 4.1.3, las diferencias divididas finitas centradas son mejores aproximaciones de las derivadas, que las versiones hacia adelante o hacia atrás. Una aproximación centrada, como la de la ecuación (25.26) tiene un error de truncamiento local de $O(h^2)$ en comparación con la aproximación hacia adelante del método de Euler, que tiene un error de $O(h)$. En consecuencia, los errores local y global del método del punto medio son $O(h^3)$ y $O(h^2)$, respectivamente.

25.2.3 Algoritmos computacionales para los métodos de Heun y del punto medio

Ambos métodos, el de Heun con un solo corrector y el del punto medio, se programan fácilmente con la estructura general mostrada en la figura 25.7. Como en las figuras 25.13*a* y 25.13*b*, es posible escribir rutinas simples en lugar de la rutina de Euler en la figura 25.7.

No obstante, cuando se va a implementar la versión iterativa del método de Heun, las modificaciones son un poco más complicadas (aunque se localicen dentro de un solo módulo). Hemos desarrollado el pseudocódigo para este propósito en la figura 25.13*c*. Este algoritmo se combina con la figura 25.7 con el objetivo de desarrollar el software para el método iterativo de Heun.

25.2.4 Resumen

Al mejorar el método de Euler desarrollamos dos nuevas técnicas de segundo orden. Aun cuando esas versiones requieren más cálculos para determinar la pendiente, la reducción que se obtiene del error nos permitirá concluir, en una sección próxima (sección 25.3.4), que usualmente una mejor exactitud vale el esfuerzo. Aunque existen ciertos casos donde técnicas fácilmente programables, como el método de Euler, pueden aplicarse con ventaja, los métodos de Heun y del punto medio por lo común son superiores y se deberán implementar si son consistentes con los objetivos del problema.

a) Heun simple sin corrector

```
SUB Heun (x, y, h, ynew)
  CALL Derivs (x, y, dy1dx)
  ye = y + dy1dx · h
  CALL Derivs(x + h, ye, dy2dx)
  Slope = (dy1dx + dy2dx)/2
  ynew = y + Slope · h
  x = x + h
END SUB
```

b) Método del punto medio

```
SUB Midpoint (x, y, h, ynew)
  CALL Derivs(x, y, dydx)
  ym = y + dydx · h/2
  CALL Derivs (x + h/2, ym, dymdx)
  ynew = y + dymdx · h
  x = x + h
END SUB
```

c) Heun con corrector

```
SUB HeunIter (x, y, h, ynew)
  es = 0.01
  maxit = 20
  CALL Derivs(x, y, dy1dx)
  ye = y + dy1dx · h
  iter = 0
  DO
    yeold = ye
    CALL Derivs(x + h, ye, dy2dx)
    slope = (dy1dx + dy2dx)/2
    ye = y + slope · h
    iter = iter + 1
    ea = |(ye - yeold)/ye| 100%
    IF (ea ≤ es OR iter > maxit) EXIT
  END DO
  ynew = ye
  x = x + h
END SUB
```

Figura 25.13
Pseudocódigo para implementar los métodos de *a*) Heun simple, *b*) punto medio y *c*) Heun iterativo.

Como se hace notar al inicio de esta sección, los métodos de Heun (sin iteraciones), del punto medio y, de hecho, la técnica de Euler misma son versiones de una clase más amplia de procedimientos de un paso denominada métodos de Runge-Kutta. Ahora veremos el desarrollo formal de esas técnicas.

25.3 MÉTODOS DE RUNGE-KUTTA

Los métodos de *Runge-Kutta* (*RK*) logran la exactitud del procedimiento de la serie de Taylor sin necesitar el cálculo de derivadas de orden superior. Existen muchas variantes, pero todas tienen la forma generalizada de la ecuación (25.1):

$$y_{i+1} = y_i + \phi(x_i, y_i, h)h \tag{25.28}$$

donde $\phi(x_i, y_i, h)$ se conoce como *función incremento*, la cual puede interpretarse como una pendiente representativa en el intervalo. La función incremento se escribe en forma general como

$$\phi = a_1k_1 + a_2k_2 + \cdots + a_nk_n \tag{25.29}$$

donde las a son constantes y las k son

$$k_1 = f(x_i, y_i) \tag{25.29a}$$

$$k_2 = f(x_i + p_1h, y_i + q_{11}k_1h) \tag{25.29b}$$

$$k_3 = f(x_i + p_2h, y_i + q_{21}k_1h + q_{22}k_2h) \tag{25.29c}$$

$$\vdots$$

$$k_n = f(x_i + p_{n-1}h, y_i + q_{n-1,1}k_1h + q_{n-1,2}k_2h + \cdots + q_{n-1,n-1}k_{n-1}h) \tag{25.29d}$$

donde las p y las q son constantes. Observe que las k son relaciones de recurrencia. Es decir, k_1 aparece en la ecuación k_2, la cual aparece en la ecuación k_3, etc. Como cada k es una evaluación funcional, esta recurrencia vuelve eficientes a los métodos *RK* para cálculos en computadora.

Es posible tener varios tipos de métodos de Runge-Kutta empleando diferentes números de términos en la función incremento especificada por n. Observe que el método de Runge-Kutta (RK) de primer orden con $n = 1$ es, de hecho, el método de Euler. Una vez que se elige n, se evalúan las a, p y q igualando la ecuación (25.28) a los términos en la expansión de la serie de Taylor (cuadro

25.1). Así, al menos para las versiones de orden inferior, el número de términos, n, por lo común representa el orden de la aproximación. Por ejemplo, en la siguiente sección, los métodos RK de segundo orden usan la función incremento con dos términos ($n = 2$). Esos métodos de segundo orden serán exactos si la solución de la ecuación diferencial es cuadrática. Además, como los términos con h^3 y mayores se eliminan durante la deducción, el error de truncamiento local es $O(h^3)$ y el global es $O(h^2)$. En secciones subsecuentes desarrollaremos los métodos RK de tercer y cuarto órdenes ($n = 3$ y 4, respectivamente). En tales casos, los errores de truncamiento global son $O(h^3)$ y $O(h^4)$, respectivamente.

25.3.1 Métodos de Runge-Kutta de segundo orden

La versión de segundo orden de la ecuación (25.28) es

$$y_{i+1} = y_i + (a_1k_1 + a_2k_2)h \tag{25.30}$$

donde:

$$k_1 = f(x_i, y_i) \tag{25.30a}$$

$$k_2 = f(x_i + p_1h, y_i + q_{11}k_1h) \tag{25.30b}$$

Como se describe en el cuadro 25.1, los valores de a_1, a_2, p_1 y q_{11} se evalúan al igualar la ecuación (25.30) con la expansión de la serie de Taylor hasta el término de segundo orden. Al hacerlo, desarrollamos tres ecuaciones para evaluar las cuatro constantes desconocidas. Las tres ecuaciones son:

$$a_1 + a_2 = 1 \tag{25.31}$$

$$a_2p_1 = \frac{1}{2} \tag{25.32}$$

$$a_2q_{11} = \frac{1}{2} \tag{25.33}$$

Como tenemos tres ecuaciones con cuatro incógnitas, debemos dar el valor de una de estas incógnitas para determinar las otras tres. Suponga que damos un valor para a_2. Entonces se resuelven de manera simultánea las ecuaciones (25.31) a (25.33) obteniendo:

$$a_1 = 1 - a_2 \tag{25.34}$$

$$p_1 = q_{11} = \frac{1}{2a_2} \tag{25.35}$$

Debido a que podemos elegir un número infinito de valores para a_2, hay un número infinito de métodos RK de segundo orden. Cada versión daría exactamente los mismos resultados si la solución de la EDO fuera cuadrática, lineal o una constante. Sin embargo, se obtienen diferentes resultados cuando (como típicamente es el caso) la solución es más complicada. A continuación presentamos tres de las versiones más comúnmente usadas y preferidas:

Método de Heun con un solo corrector ($a_2 = 1/2$) Si suponemos que a_2 es 1/2 de las ecuaciones (25.34) y (25.35) puede obtenerse $a_1 = 1/2$ y $p_1 = q_{11} = 1$. Estos parámetros, al sustituirse en la ecuación (25.30), dan:

$$y_{i+1} = y_i + \left(\frac{1}{2}k_1 + \frac{1}{2}k_2\right)h \tag{25.36}$$

donde

$$k_1 = f(x_i, y_i) \tag{25.36a}$$

$$k_2 = f(x_i + h, y_i + k_1h) \tag{25.36b}$$

Cuadro 25.1 Deducción de los métodos de Runge-Kutta de segundo orden

La versión de segundo orden de la ecuación (25.28) es:

$$y_{i+1} = y_i + (a_1k_1 + a_2k_2)h \tag{C25.1.1}$$

donde

$$k_1 = f(x_i, y_i) \tag{C25.1.2}$$

y

$$k_2 = f(x_i + p_1h, y_i + q_{11}k_1h) \tag{C25.1.3}$$

Para usar la ecuación (C25.1.1) debemos determinar los valores de las constantes a_1, a_2, p_1 y q_{11}. Para ello, recordamos que la serie de Taylor de segundo orden para y_{i+1}, en términos de y_i y $f(x_i, y_i)$, se escribe como [ecuación (25.11)]:

$$y_{i+1} = y_i + f(x_i, y_i)h + \frac{f'(x_i, y_i)}{2!}h^2 \tag{C25.1.4}$$

donde $f(x_i, y_i)$ debe determinarse por derivación usando la regla de la cadena (sección 25.1.3):

$$f'(x_i, y_i) = \frac{\partial f(x, y)}{\partial x} + \frac{\partial f(x, y)}{\partial y}\frac{dy}{dx} \tag{C25.1.5}$$

Si sustituimos la ecuación (C25.1.5) en la ecuación (C25.1.4) se obtiene:

$$y_{i+1} = y_i + f(x_i, y_i)h + \left(\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx}\right)\frac{h^2}{2!} \tag{C25.1.6}$$

La estrategia básica de los métodos de Runge-Kutta es el uso de manipulaciones algebraicas para obtener los valores de a_1, a_2, p_1 y q_{11}, que hacen equivalentes a las ecuaciones (C25.1.1) y (C25.1.6).

Para ello, primero usamos una serie de Taylor para expandir la ecuación (C25.1.3). La serie de Taylor para una función de dos variables se define como [recuerde la ecuación (4.26)]:

$$g(x + r, y + s) = g(x, y) + r\frac{\partial g}{\partial x} + s\frac{\partial g}{\partial y} + \cdots$$

Si se aplica este método para expandir la ecuación (C25.1.3) se llega a:

$$f(x_i + p_1h, y_i + q_{11}k_1h) = f(x_i, y_i) + p_1h\frac{\partial f}{\partial x} + q_{11}k_1h\frac{\partial f}{\partial y} + O(h^2)$$

Este resultado se sustituye junto con la ecuación (C25.1.2) en la ecuación (C25.1.1) para obtener:

$$y_{i+1} = y_i + a_1hf(x_i, y_i) + a_2hf(x_i, y_i) + a_2p_1h^2\frac{\partial f}{\partial x} + a_2q_{11}h^2f(x_i, y_i)\frac{\partial f}{\partial y} + O(h^3)$$

o, agrupando términos,

$$y_{i+1} = y_i + [a_1f(x_i, y_i) + a_2f(x_i, y_i)]h + \left[a_2p_1\frac{\partial f}{\partial x} + a_2q_{11}f(x_i, y_i)\frac{\partial f}{\partial y}\right]h^2 + O(h^3) \tag{C25.1.7}$$

Ahora, si comparamos términos comunes en las ecuaciones (C25.1.6) y (C25.1.7), determinamos que para que las dos ecuaciones sean equivalentes, se debe satisfacer lo siguiente:

$$a_1 + a_2 = 1$$

$$a_2p_1 = \frac{1}{2}$$

$$a_2q_{11} = \frac{1}{2}$$

Las tres ecuaciones simultáneas anteriores contienen las cuatro constantes desconocidas. Como hay una incógnita más que el número de ecuaciones, no existe un conjunto único de constantes que satisfaga las ecuaciones. Sin embargo, considerando un valor para una de las constantes, es posible determinar el valor de las otras tres. En consecuencia, existe una familia de métodos de segundo orden y no una sola versión.

Observe que k_1 es la pendiente al inicio del intervalo y que k_2 es la pendiente al final del intervalo. En consecuencia, este método de Runge-Kutta de segundo orden es, de hecho, la técnica de Heun sin iteración.

El método del punto medio ($a_2 = 1$) Si suponemos que a_2 es 1, entonces $a_1 = 0$, $p_1 = q_{11} = 1/2$, y la ecuación (25.30) se convierte en:

$$y_{i+1} = y_i + k_2h \tag{25.37}$$

donde

$$k_1 = f(x_i, y_i) \tag{25.37a}$$

$$k_2 = f\left(x_i + \frac{1}{2}h, y_i + \frac{1}{2}k_1 h\right) \qquad (25.37b)$$

Éste es el método del punto medio.

Método de Ralston (a_2 = 2/3) Ralston (1962) y Ralston y Rabinowitz (1978) determinaron que al seleccionar $a_2 = 2/3$ se obtiene un mínimo en el error de truncamiento para los algoritmos RK de segundo orden. Con esta versión, $a_1 = 1/3$ y $p_1 = q_{11} = 3/4$ y da:

$$y_{i+1} = y_i + \left(\frac{1}{3}k_1 + \frac{2}{3}k_2\right)h \qquad (25.38)$$

donde

$$k_1 = f(x_i, y_i) \qquad (25.38a)$$

$$k_2 = f\left(x_i + \frac{3}{4}h, y_i + \frac{3}{4}k_1 h\right) \qquad (25.38b)$$

EJEMPLO 25.6 Comparación de varios esquemas RK de segundo orden

Planteamiento del problema Utilice los métodos de punto medio [ecuación (25.37)] y el de Ralston [ecuación (25.38)] para integrar numéricamente la ecuación (PT7.13):

$$f(x, y) = -2x^3 + 12x^2 - 20x + 8.5$$

desde $x = 0$ hasta $x = 4$, usando un tamaño de paso de 0.5. La condición inicial es $x = 0$, $y = 1$. Compare los resultados con los valores obtenidos usando otro algoritmo RK de segundo orden: el método de Heun sin iteración del corrector (tabla 25.3).

Tabla 25.3 Comparación de los valores verdadero y aproximado de la integral de $y' = -2x^3 + 12x^2 - 20x + 8.5$, con la condición inicial de que $y = 1$ en $x = 0$. Los valores aproximados se calcularon por medio de tres versiones de los métodos RK de segundo orden, con un tamaño de paso de 0.5.

		Heun		Punto medio		RK Ralston de segundo orden	
x	$y_{verdadero}$	y	$\|\varepsilon_t\|$(%)	y	$\|\varepsilon_t\|$(%)	y	$\|\varepsilon_t\|$(%)
0.0	1.00000	1.00000	0	1.00000	0	1.00000	0
0.5	3.21875	3.43750	6.8	3.109375	3.4	3.277344	1.8
1.0	3.00000	3.37500	12.5	2.81250	6.3	3.101563	3.4
1.5	2.21875	2.68750	21.1	1.984375	10.6	2.347656	5.8
2.0	2.00000	2.50000	25.0	1.75	12.5	2.140625	7.0
2.5	2.71875	3.18750	17.2	2.484375	8.6	2.855469	5.0
3.0	4.00000	4.37500	9.4	3.81250	4.7	4.117188	2.9
3.5	4.71875	4.93750	4.6	4.609375	2.3	4.800781	1.7
4.0	3.00000	3.00000	0	3	0	3.031250	1.0

Solución El primer paso en el método de punto medio consiste en usar la ecuación (25.37*a*) para calcular:

$$k_1 = -2(0)^3 + 12(0)^2 - 20(0) + 8.5 = 8.5$$

Sin embargo, como la EDO está en función sólo de x, este resultado carece de relevancia sobre el segundo paso [el uso de la ecuación (25.37*b*)] para calcular:

$$k_2 = -2(0.25)^3 + 12(0.25)^2 - 20(0.25) + 8.5 = 4.21875$$

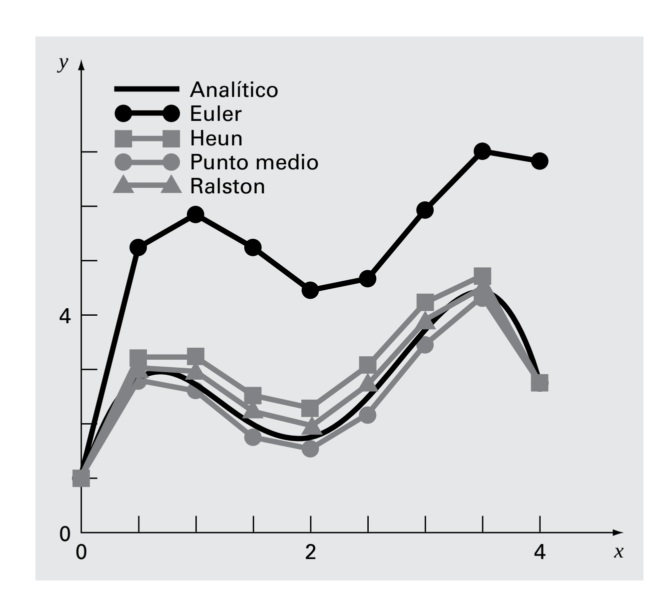


Figura 25.14
Comparación de la solución verdadera con soluciones numéricas usando tres métodos RK de segundo orden y el método de Euler.

Observe que tal estimación de la pendiente es mucho más cercana al valor promedio en el intervalo (4.4375), que la pendiente al inicio del intervalo (8.5) que se habría usado con el procedimiento de Euler. La pendiente en el punto medio entonces se sustituye en la ecuación (25.37) para predecir:

$$y(0.5) = 1 + 4.21875(0.5) = 3.109375 \qquad \varepsilon_t = 3.4\%$$

El cálculo se repite; los resultados se resumen en la figura 25.14 y en la tabla 25.3.

En el método de Ralston, k_1 en el primer intervalo también es igual a 8.5 y [ecuación (25.38*b*)]

$$k_2 = -2(0.375)^3 + 12(0.375)^2 - 20(0.375) + 8.5 = 2.58203125$$

La pendiente promedio se calcula mediante:

$$\phi = \frac{1}{3}(8.5) + \frac{2}{3}(2.58203125) = 4.5546875$$

que se utiliza para predecir:

$$y(0.5) = 1 + 4.5546875(0.5) = 3.27734375 \qquad \varepsilon_t = -1.82\%$$

Los cálculos se repiten; los resultados se resumen en la figura 25.14 y en la tabla 25.3. Observe que todos los métodos RK de segundo orden son superiores al método de Euler.

25.3.2 Métodos de Runge-Kutta de tercer orden

Para $n = 3$, es posible efectuar un desarrollo similar al del método de segundo orden. El resultado de tal desarrollo genera seis ecuaciones con ocho incógnitas. Por lo tanto, se deben dar *a priori* los valores de dos de las incógnitas con la finalidad de establecer los parámetros restantes. Una versión común que se obtiene es

$$y_{i+1} = y_i + \frac{1}{6}(k_1 + 4k_2 + k_3)h \tag{25.39}$$

donde

$$k_1 = f(x_i, y_i) \tag{25.39a}$$

$$k_2 = f\left(x_i + \frac{1}{2}h, y_i + \frac{1}{2}k_1 h\right) \tag{25.39b}$$

$$k_3 = f(x_i + h, y_i - k_1 h + 2k_2 h) \tag{25.39c}$$

Observe que si la EDO está en función sólo de x, este método de tercer orden se reduce a la regla de Simpson 1/3. Ralston (1962) y Ralston y Rabinowitz (1978) desarrollaron una versión alternativa que proporciona un mínimo para el error de truncamiento. En cualquier caso, los métodos RK de tercer orden tienen errores local y global de $O(h^4)$ y $O(h^3)$, respectivamente, y dan resultados exactos cuando la solución es una cúbica. Al tratarse de polinomios, la ecuación (25.39) será también exacta cuando la ecuación diferencial sea cúbica y la solución sea de cuarto grado. Ello se debe a

que la regla de Simpson 1/3 ofrece estimaciones exactas de la integral para cúbicas (recuerde el cuadro 21.3).

25.3.3 Métodos de Runge-Kutta de cuarto orden

El más popular de los métodos RK es el de cuarto orden. Como en el caso de los procedimientos de segundo orden, hay un número infinito de versiones. La siguiente, es la forma comúnmente usada y, por lo tanto, le llamamos *método clásico RK de cuarto orden*:

$$y_{i+1} = y_i + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)h \tag{25.40}$$

donde

$$k_1 = f(x_i, y_i) \tag{25.40a}$$

$$k_2 = f\left(x_i + \frac{1}{2}h, y_i + \frac{1}{2}k_1 h\right) \tag{25.40b}$$

$$k_3 = f\left(x_i + \frac{1}{2}h, y_i + \frac{1}{2}k_2 h\right) \tag{25.40c}$$

$$k_4 = f(x_i + h, y_i + k_3 h) \tag{25.40d}$$

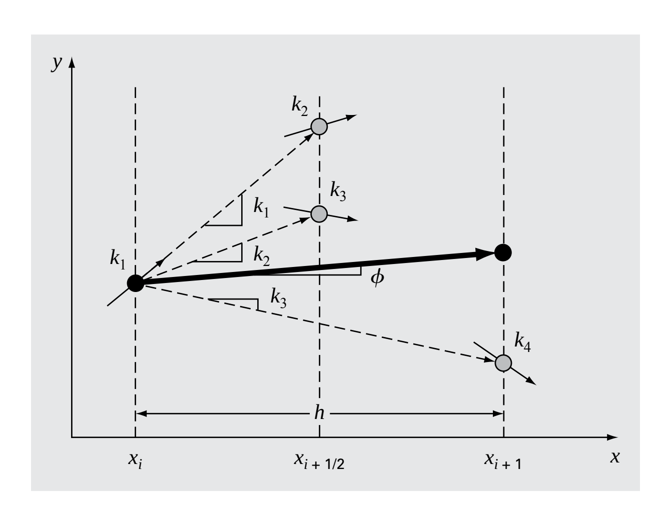


Figura 25.15
Representación gráfica de las pendientes estimadas empleadas en el método RK de cuarto orden.

Observe que con las EDO que están en función sólo de x, el método RK clásico de cuarto orden es similar a la regla de Simpson 1/3. Además, el método RK de cuarto orden tiene similitud con el procedimiento de Heun en cuanto a que se usan múltiples estimaciones de la pendiente para obtener una mejor pendiente promedio en el intervalo. Como se muestra en la figura 25.15, cada una de las k representa una pendiente. La ecuación (25.40) entonces representa un promedio ponderado de éstas para establecer la mejor pendiente.

EJEMPLO 25.7 Método clásico RK de cuarto orden

Planteamiento del problema

a) Con el método clásico RK de cuarto orden [ecuación (25.40)] integre

$$f(x, y) = -2x^3 + 12x^2 - 20x + 8.5$$

usando un tamaño de paso $h = 0.5$ y la condición inicial $y = 1$ en $x = 0$;

b) De manera similar integre

$$f(x, y) = 4e^{0.8x} - 0.5y$$

utilizando $h = 0.5$ con $y(0) = 2$ desde $x = 0$ hasta 0.5.

Solución

a) Se emplean las ecuaciones (25.40*a*) a (25.40*d*) para calcular $k_1 = 8.5$, $k_2 = 4.21875$, $k_3 = 4.21875$ y $k_4 = 1.25$; las cuales se sustituyen en la ecuación (25.40) para dar

$$y(0.5) = 1 + \left\{\frac{1}{6}[8.5 + 2(4.21875) + 2(4.21875) + 1.25]\right\} 0.5$$
$$= 3.21875$$

que es exacta. Por consiguiente, como la solución verdadera es una cuártica [ecuación (PT7.16)], el método de cuarto orden da un resultado exacto.

b) En este caso, la pendiente al inicio del intervalo se calcula como sigue:

$k_1 = f(0, 2) = 4e^{0.8(0)} - 0.5(2) = 3$

Este valor se utiliza para calcular un valor de y y una pendiente en el punto medio,

$y(0.25) = 2 + 3(0.25) = 2.75$
$k_2 = f(0.25, 2.75) = 4e^{0.8(0.25)} - 0.5(2.75) = 3.510611$

Esta pendiente, a su vez, se utiliza para calcular otro valor de y y otra pendiente en el punto medio,

$y(0.25) = 2 + 3.510611(0.25) = 2.877653$
$k_3 = f(0.25, 2.877653) = 4e^{0.8(0.25)} - 0.5(2.877653) = 3.446785$

Después, se usará esta pendiente para calcular un valor de y y una pendiente al final del intervalo,

$y(0.5) = 2 + 3.071785(0.5) = 3.723392$
$k_4 = f(0.5, 3.723392) = 4e^{0.8(0.5)} - 0.5(3.723392) = 4.105603$

Por último, las cuatro estimaciones de la pendiente se combinan para obtener una pendiente promedio, la cual se utiliza después para realizar la última predicción al final del intervalo.

$$\phi = \frac{1}{6}[3 + 2(3.510611) + 2(3.446785) + 4.105603] = 3.503399$$
$$y(0.5) = 2 + 3.503399(0.5) = 3.751669$$

que es muy aproximada a la solución verdadera de 3.751521.

25.3.4 Métodos de Runge-Kutta de orden superior

Cuando se requieren resultados más exactos, se recomienda el *método RK de quinto orden de Butcher* (1964):

$$y_{i+1} = y_i + \frac{1}{90}(7k_1 + 32k_3 + 12k_4 + 32k_5 + 7k_6)h \tag{25.41}$$

donde

$$k_1 = f(x_i, y_i) \tag{25.41a}$$

$$k_2 = f\left(x_i + \frac{1}{4}h, y_i + \frac{1}{4}k_1h\right) \tag{25.41b}$$

$$k_3 = f\left(x_i + \frac{1}{4}h, y_i + \frac{1}{8}k_1h + \frac{1}{8}k_2h\right) \tag{25.41c}$$

$$k_4 = f\left(x_i + \frac{1}{2}h, y_i - \frac{1}{2}k_2h + k_3h\right) \tag{25.41d}$$

$$k_5 = f\left(x_i + \frac{3}{4}h, y_i + \frac{3}{16}k_1h + \frac{9}{16}k_4h\right) \tag{25.41e}$$

$$k_6 = f\left(x_i + h, y_i - \frac{3}{7}k_1h + \frac{2}{7}k_2h + \frac{12}{7}k_3h - \frac{12}{7}k_4h + \frac{8}{7}k_5h\right) \tag{25.41f}$$

Observe la semejanza entre el método de Butcher y la regla de Boole de la tabla 21.2. Existen las fórmulas RK de orden superior, como el método de Butcher, pero en general, la ganancia en exactitud con métodos mayores al cuarto orden se ve afectada por mayor trabajo computacional y mayor complejidad.

EJEMPLO 25.8 Comparación de los métodos de Runge-Kutta

Planteamiento del problema Con los métodos RK desde primero hasta quinto orden resuelva

$$f(x, y) = 4e^{0.8x} - 0.5y$$

con $y(0) = 2$ desde $x = 0$ hasta $x = 4$ con diferentes tamaños de paso. Compare la exactitud de los diferentes métodos para la estimación en $x = 4$, basándose en la respuesta exacta a cinco decimales, $y(4) = 75.33896$.

Solución El cálculo se realiza usando los métodos de Euler, de Heun no iterativo, RK de tercer orden [ecuación (25.39)], clásico RK de cuarto orden y RK de quinto orden de Butcher. Los resultados se presentan en la figura 25.16, donde graficamos el valor absoluto del error relativo porcentual contra el trabajo computacional. Esta última cantidad es equivalente al número requerido de evaluaciones de la función para obtener el resultado, como:

$$\text{Trabajo computacional} = n_f \frac{b-a}{h} \qquad \text{(E25.8.1)}$$

donde n_f = número de evaluaciones de la función consideradas para el cálculo particular de RK. Para órdenes ≤ 4, n_f es igual al orden del método; sin embargo, observe que la técnica de Butcher de quinto orden requiere seis evaluaciones de la función [ecuaciones (25.41*a*) a la (25.41*f*)]. La cantidad $(b - a)/h$ es el intervalo de integración total dividido entre el tamaño de paso (es decir, es el número necesario de aplicaciones de la técnica RK para obtener el resultado. Como las evaluaciones de la función son generalmente las que consumen más tiempo, la ecuación (E25.8.1) proporciona una medida burda del tiempo de ejecución requerido para obtener la respuesta.

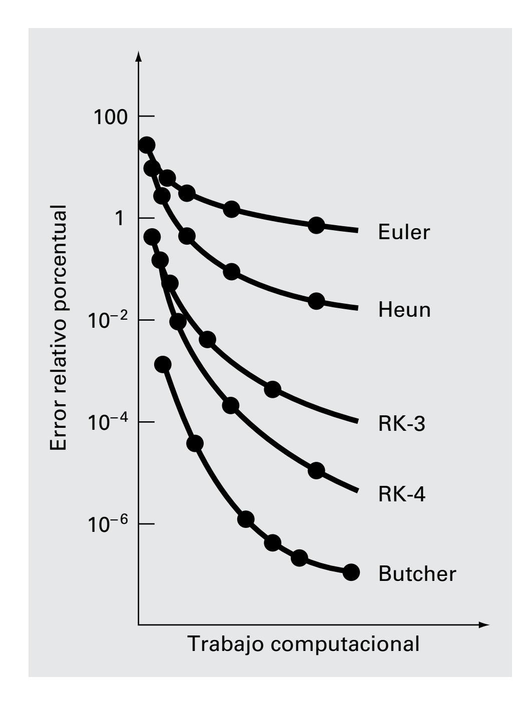


Figura 25.16
Comparación del error relativo porcentual contra trabajo computacional para los métodos de RK, desde el de primero hasta el quinto órdenes.

La inspección de la figura 25.16 nos lleva a varias conclusiones: primero, que los métodos de orden superior logran mayor exactitud con el mismo trabajo computacional; segundo, que la ganancia en exactitud lograda por el mayor trabajo tiende a disminuir después de un punto. (Observe que las curvas primero caen con rapidez y después tienden a nivelarse.)

```
SUB RK4 (x, y, h, ynew)
  CALL Derivs(x, y, k1)
  ym = y + k1 · h/2
  CALL Derivs(x + h/2, ym, k2)
  ym = y + k2 · h/2
  CALL Derivs(x + h/2, ym, k3)
  ye = y + k3 · h
  CALL Derivs(x + h, ye, k4)
  slope = (k1 + 2(k2 + k3) + k4)/6
  ynew = y + slope · h
  x = x + h
END SUB
```

Figura 25.17
Pseudocódigo para determinar un solo paso del método RK de cuarto orden.

El ejemplo 25.8 y la figura 25.16 nos llevarán a la conclusión de que las técnicas RK de orden superior son siempre los métodos de preferencia. Sin embargo, deben considerarse otros factores cuando se elija una técnica de solución, como el costo de programación y los requerimientos de exactitud del problema. Las alternativas (ventajas y desventajas) se explorarán con detalle en las aplicaciones a la ingeniería en el capítulo 28 y en el epílogo de la parte siete.

25.3.5 Algoritmos computacionales para los métodos de Runge-Kutta

Como en el caso de todos los métodos expuestos en este capítulo, las técnicas RK se ajustan muy bien al algoritmo general formulado en la figura 25.7. La figura 25.17 muestra el pseudocódigo para determinar la pendiente del método clásico RK de cuarto orden [ecuación (25.40)]. Las subrutinas que calculan las pendientes para todas las otras versiones se programan fácilmente de forma similar.

25.4 SISTEMAS DE ECUACIONES

Muchos problemas prácticos en la ingeniería y en la ciencia requieren la solución de un sistema de ecuaciones diferenciales ordinarias simultáneas más que de una sola ecuación. Tales sistemas en general se representan como:

$$\frac{dy_1}{dx} = f_1(x, y_1, y_2, \ldots, y_n)$$
$$\frac{dy_2}{dx} = f_2(x, y_1, y_2, \ldots, y_n)$$
$$\vdots$$
$$\frac{dy_n}{dx} = f_n(x, y_1, y_2, \ldots, y_n) \tag{25.42}$$

La solución de este sistema requiere que se conozcan n condiciones iniciales en el valor inicial de x.

25.4.1 Método de Euler

Todos los métodos analizados en este capítulo, para ecuaciones solas, pueden extenderse al sistema que se mostró antes. Las aplicaciones en la ingeniería llegan a considerar miles de ecuaciones simultáneas. En todo caso, el procedimiento para resolver un sistema de ecuaciones consiste únicamente en aplicar la técnica simple por ecuación en cada paso, antes de proceder con el siguiente. Lo anterior se ilustra mejor con el siguiente ejemplo para el método de Euler simple.

EJEMPLO 25.9 Solución de sistemas de EDO usando el método de Euler

Planteamiento del problema Resuelva el siguiente sistema de ecuaciones diferenciales utilizando el método de Euler, suponiendo que en $x = 0$, $y_1 = 4$ y $y_2 = 6$. Integre hasta $x = 2$ con un tamaño de paso igual a 0.5.

$$\frac{dy_1}{dx} = -0.5y_1 \qquad \frac{dy_2}{dx} = 4 - 0.3y_2 - 0.1y_1$$

Solución Se implementa el método de Euler para cada variable como en la ecuación (25.2):

$$y_1(0.5) = 4 + [-0.5(4)]0.5 = 3$$
$$y_2(0.5) = 6 + [4 - 0.3(6) - 0.1(4)]0.5 = 6.9$$

Observe que $y_1(0) = 4$ se emplea en la segunda ecuación en lugar de $y_1(0.5) = 3$ calculada con la primera ecuación. Procediendo de manera similar se tiene:

x	y_1	y_2
0	4	6
0.5	3	6.9
1.0	2.25	7.715
1.5	1.6875	8.44525
2.0	1.265625	9.094087

25.4.2 Métodos de Runge-Kutta

Observe que cualquiera de los métodos RK de orden superior expuestos en este capítulo se pueden aplicar a los sistemas de ecuaciones. Sin embargo, debe tenerse cuidado al determinar las pendientes. La figura 25.15 es útil para visualizar la forma adecuada de hacer esto con el método de cuarto orden. Es decir, desarrollamos primero las pendientes para todas las variables en el valor inicial. Esas pendientes (un conjunto de las k_1) se utilizarán después para realizar predicciones de la variable dependiente en el punto medio del intervalo. Tales valores del punto medio se utilizan, a su vez, para calcular un conjunto de pendientes en el punto medio (las k_2). Esas nuevas pendientes se vuelven a usar en el punto de inicio para efectuar otro conjunto de predicciones del punto medio que lleven a nuevas predicciones de la pendiente en el punto medio (las k_3). Éstas después se emplearán para realizar predicciones al final del intervalo que se usarán para desarrollar pendientes al final del intervalo (las k_4). Por último, las k se combinan en un conjunto de funciones incrementadas [como en la ecuación (25.40)] y se llevan de nuevo al inicio para hacer la predicción final. El siguiente ejemplo ilustra el procedimiento.

EJEMPLO 25.10 Solución de sistemas de EDO usando el método RK de cuarto orden

Planteamiento del problema Con el método RK de cuarto orden resuelva las EDO del ejemplo 25.9.

Solución Primero debemos encontrar todas las pendientes al inicio del intervalo:

$$k_{1,1} = f_1(0, 4, 6) = -0.5(4) = -2$$
$$k_{1,2} = f_2(0, 4, 6) = 4 - 0.3(6) - 0.1(4) = 1.8$$

donde $k_{i,j}$ es el i-ésimo valor de k para la j-ésima variable dependiente. Después, se requiere calcular los primeros valores de y_1 y y_2 en el punto medio:

$$y_1 + k_{1,1}\frac{h}{2} = 4 + (-2)\frac{0.5}{2} = 3.5$$
$$y_2 + k_{1,2}\frac{h}{2} = 6 + (1.8)\frac{0.5}{2} = 6.45$$

que se utilizarán para calcular el primer conjunto de pendientes en el punto medio,

$$k_{2,1} = f_1(0.25, 3.5, 6.45) = -1.75$$
$$k_{2,2} = f_2(0.25, 3.5, 6.45) = 1.715$$

Éstas sirven para determinar el segundo conjunto de predicciones en el punto medio,

$$y_1 + k_{2,1}\frac{h}{2} = 4 + (-1.75)\frac{0.5}{2} = 3.5625$$
$$y_2 + k_{2,2}\frac{h}{2} = 6 + (1.715)\frac{0.5}{2} = 6.42875$$

que se usan para calcular el segundo conjunto de pendientes en el punto medio,

$$k_{3,1} = f_1(0.25, 3.5625, 6.42875) = -1.78125$$
$$k_{3,2} = f_2(0.25, 3.5625, 6.42875) = 1.715125$$

Éstas se utilizarán para determinar las predicciones al final del intervalo:

$$y_1 + k_{3,1}h = 4 + (-1.78125)(0.5) = 3.109375$$
$$y_2 + k_{3,2}h = 6 + (1.715125)(0.5) = 6.857563$$

que se usan para calcular las pendientes al final del intervalo:

$k_{4,1} = f_1(0.5, 3.109375, 6.857563) = -1.554688$

$k_{4,2} = f_2(0.5, 3.109375, 6.857563) = 1.631794$

Los valores de k se utilizan después para calcular [ecuación (25.40)]:

$$y_1(0.5) = 4 + \frac{1}{6}[-2 + 2(-1.75 - 1.78125) - 1.554688]0.5 = 3.115234$$

$$y_2(0.5) = 6 + \frac{1}{6}[1.8 + 2(1.715 + 1.715125) + 1.631794]0.5 = 6.857670$$

Procediendo de la misma forma con los pasos restantes se obtiene

x	y_1	y_2
0	4	6
0.5	3.115234	6.857670
1.0	2.426171	7.632106
1.5	1.889523	8.326886
2.0	1.471577	8.946865

25.4.3 Algoritmo computacional para resolver sistemas de EDO

El código computacional para resolver una sola EDO con el método de Euler (figura 25.7) puede fácilmente extenderse a sistemas de ecuaciones. Las modificaciones son:

1. Dar el número de ecuaciones, n.
2. Dar los valores iniciales para cada una de las n variables dependientes.
3. Modificar el algoritmo de manera que calcule las pendientes para cada una de las variables dependientes.
4. Incluir las ecuaciones adicionales para calcular los valores de la derivada por cada una de las EDO.
5. Incluir ciclos para calcular un nuevo valor para cada variable dependiente.

Este algoritmo se presenta en la figura 25.18 para el método RK de cuarto orden. Observe que es similar en estructura y organización a la figura 25.7. La mayoría de las diferencias se relacionan con el hecho de que:

1. Hay n ecuaciones.
2. Está el detalle adicionado del método RK de cuarto orden.

EJEMPLO 25.11 Solución de sistemas de EDO con el uso de computadora

Planteamiento del problema Un programa de cómputo para implementar el método RK de cuarto orden para sistemas se puede desarrollar fácilmente con base en las figuras 25.18. Tal software es conveniente para comparar diferentes modelos de un sistema físico. Por ejemplo, un modelo lineal para un péndulo oscilante está dado por [recuerde la ecuación (PT7.11)]:

$$\frac{dy_1}{dx} = y_2 \qquad \frac{dy_2}{dx} = -16.1y_1$$

donde y_1 y y_2 = desplazamiento angular y velocidad. Un modelo no lineal del mismo sistema es [recuerde la ecuación (PT7.9)]:

***a)* Programa principal o “manejador”**

```
Asigna valores para
n = número de ecuaciones
yi = valores iniciales de n variables
     dependientes
xi = valor inicial de la variable
     independiente
xf = valor final de la variable
     independiente
dx = cálculo del tamaño de paso
xout = intervalo de salida

x = xi
m = 0
xpm = x
DOFOR i = 1, n
  ypi,m = yii
  yi = yii
END DO
DO
  xend = x + xout
  IF (xend > xf) THEN xend = xf
  h = dx
  CALL Integrator (x, y, n, h, xend)
  m = m + 1
  xpm = x
  DOFOR i = 1, n
    ypi,m = yi
  END DO
  IF (x ≥ xf) EXIT
END DO
DISPLAY RESULTS
END
```

***b)* Rutina para tomar un paso de salida**

```
SUB Integrator (x, y, n, h, xend)
  DO
    IF (xend - x < h) THEN h = xend - x
    CALL RK4 (x, y, n, h)
    IF (x ≥ xend) EXIT
  END DO
END SUB
```

***c)* Método RK de cuarto orden para un sistema de EDO**

```
SUB RK4(x, y, n, h)
  CALL Derivs (x, y, k1)
  DOFOR i = 1, n
    ymi = yi + k1i * h/2
  END DO
  CALL Derivs (x + h / 2, ym, k2)
  DOFOR i = 1, n
    ymi = yi + k2i * h / 2
  END DO
  CALL Derivs (x + h / 2, ym, k3)
  DOFOR i = 1, n
    yei = yi + k3i * h
  END DO
  CALL Derivs (x + h, ye, k4)
  DOFOR i = 1, n
    slopei = (k1i + 2*(k2i+k3i)+k4i)/6
    yi = yi + slopei * h
  END DO
  x = x + h
END SUB
```

***d)* Rutina para determinar derivadas**

```
SUB Derivs (x, y, dy)
  dy1 = ...
  dy2 = ...
END SUB
```

Figura 25.18
Pseudocódigo del método RK de cuarto orden para sistemas de ecuaciones.

$$\frac{dy_3}{dx} = y_4 \qquad \frac{dy_4}{dx} \approx -16.1 \operatorname{sen}(y_3)$$

donde y_3 y y_4 = desplazamiento angular y velocidad en el caso no lineal. Resuelva estos sistemas en dos casos: *a*) un pequeño desplazamiento inicial ($y_1 = y_3 = 0.1$ radianes; $y_2 = y_4 = 0$) y *b*) un gran desplazamiento ($y_1 = y_3 \approx \pi/4 = 0.785398$ radianes; $y_2 = y_4 = 0$).

Solución

a) Los resultados calculados para los modelos lineal y no lineal son casi idénticos (figura 25.19*a*). Esto era lo que se esperaba, ya que cuando el desplazamiento inicial es pequeño, sen $(\theta) \cong \theta$.

b) Cuando el desplazamiento inicial es $\pi/4 = 0.785398$, las soluciones son diferentes. Esta diferencia se magnifica conforme el tiempo va aumentando (figura 25.19*b*). Esto se esperaba, ya que la suposición de que sen $(\theta) = \theta$ es mala cuando theta es grande.

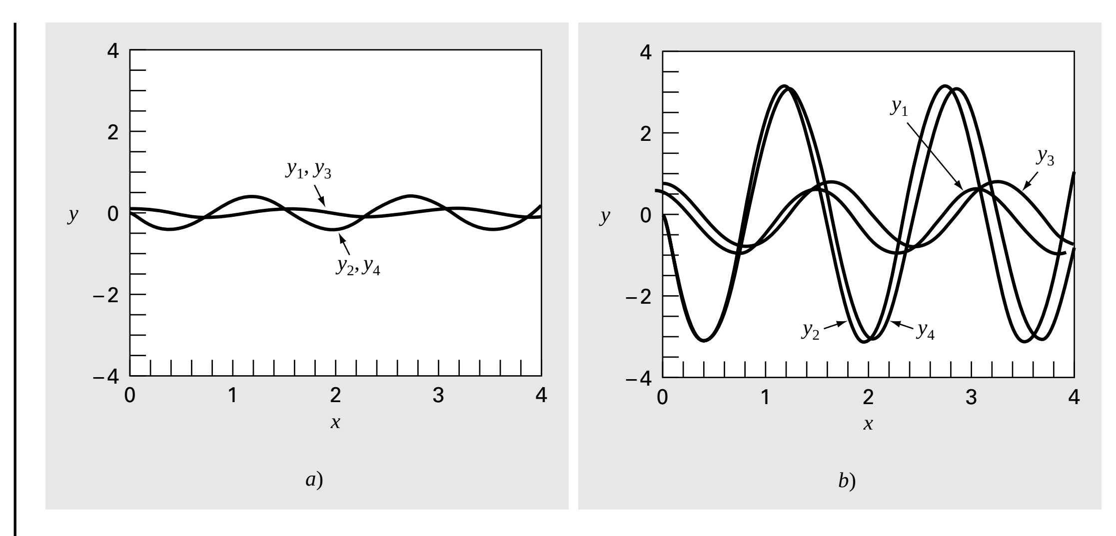


Figura 25.19
Soluciones obtenidas con un programa computacional para el método RK de cuarto orden. Las gráficas representan soluciones para péndulos tanto lineales como no lineales con desplazamientos iniciales *a*) pequeñas y *b*) grandes.

25.5 MÉTODOS ADAPTATIVOS DE RUNGE-KUTTA

Hasta ahora, se han presentado métodos para resolver las EDO que emplean un tamaño de paso constante. En un número significativo de problemas, esto llega a representar una seria limitación. Por ejemplo, suponga que pretendemos integrar una EDO con una solución del tipo expuesto en la figura 25.20. En la mayor parte del intervalo, la solución cambia de manera gradual. Tal comportamiento sugiere la posibilidad de emplear un tamaño de paso grande para obtener resultados adecuados; sin embargo, en una región localizada desde $x = 1.75$ hasta $x = 2.25$, la solución tiene un cambio abrupto. La consecuencia práctica cuando se trabaja con estas funciones es que se requeriría un tamaño de paso muy pequeño para captar en forma más precisa el comportamiento impulsivo. Si se empleara un algoritmo con tamaño de paso constante, el tamaño de paso más pequeño necesario para la región del cambio abrupto se aplicaría en todo el cálculo. En consecuencia, un tamaño de paso más pequeño que el necesario (y, por lo tanto, la implicación de más cálculos) se desperdiciaría en las regiones del cambio gradual.

Los algoritmos que ajustan automáticamente el tamaño de paso pueden evitar tal desperdicio y lograr así una gran ventaja. Como estos algoritmos se "adaptan" a la trayectoria de la solución, se dice que tienen *control adaptativo del tamaño de paso*. La implementación de tales procedimientos requiere la obtención de un estimado del error de truncamiento local en cada paso. Dicho error estimado puede servir después como base para aumentar o disminuir el tamaño de paso.

Antes de proceder al desarrollo, debemos mencionar que además de resolver las EDO, los métodos descritos en este capítulo también se utilizan para evaluar integrales definidas. Como se menciona en la introducción de la parte seis, la evaluación de la integral:

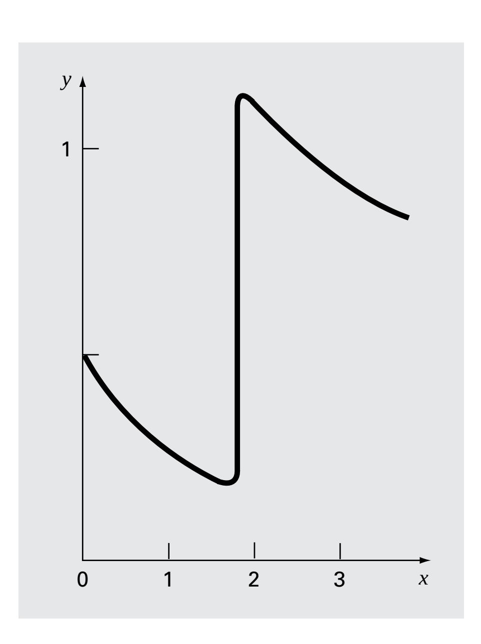


Figura 25.20
Ejemplo de la solución para una EDO que exhibe un cambio abrupto. El ajuste automático del tamaño de paso tiene grandes ventajas en esos casos.

$$I = \int_a^b f(x)\, dx$$

es equivalente a resolver la ecuación diferencial:

$$\frac{dy}{dx} = f(x)$$

para $y(b)$ dada la condición inicial $y(a) = 0$. Así, las siguientes técnicas se emplean para evaluar con eficiencia las integrales definidas de funciones que, en general, son suaves, pero que exhiben regiones de cambio abrupto.

Existen dos procedimientos importantes para incorporar el control adaptativo del tamaño de paso en los métodos de un paso. En el primero, el error se estima como la diferencia entre dos predicciones usando el método RK del mismo orden, aunque con diferentes tamaños de paso. En el segundo, el error de truncamiento local se estima como la diferencia entre dos predicciones usando métodos RK de diferente orden.

25.5.1 Método adaptativo de RK o de mitad de paso

El método de *mitad de paso* (o *adaptativo de RK*) consiste en realizar dos veces cada paso, una vez como un solo paso e, independientemente, como dos medios pasos. La diferencia entre los dos resultados representa un estimado del error de truncamiento local. Si y_1 representa la predicción con un solo paso, y y_2, la predicción con dos medios pasos, el error Δ se representa como:

$$\Delta = y_2 - y_1 \qquad (25.43)$$

Además de proporcionar un criterio para el control del tamaño de paso, la ecuación (25.43) también se utiliza para corregir la predicción y_2. En la versión RK de cuarto orden, la corrección es:

$$y_2 \leftarrow y_2 + \frac{\Delta}{15} \qquad (25.44)$$

Dicha estimación tiene una precisión de quinto orden.

EJEMPLO 25.12 Método adaptativo de RK de cuarto orden

Planteamiento del problema Utilice el método adaptativo de RK de cuarto orden para integrar $y' = 4e^{0.8x} - 0.5y$ desde $x = 0$ hasta 2 usando $h = 2$ y la condición inicial $y(0) = 2$. Ésta es la misma ecuación diferencial que se resolvió antes en el ejemplo 25.5. Recuerde que la solución verdadera es $y(2) \approx 14.84392$.

Solución La predicción sencilla con un tamaño de paso h se calcula como sigue:

$$y(2) = 2 + \frac{1}{6}[3 + 2(6.40216 + 4.70108) + 14.11105]2 = 15.10584$$

Las dos predicciones de medio paso son:

$$y(1) = 2 + \frac{1}{6}[3 + 2(4.21730 + 3.91297) + 5.945681]1 = 6.20104$$

y

$$y(2) = 6.20104 + \frac{1}{6}[5.80164 + 2(8.72954 + 7.99756) + 12.71283]1 = 14.86249$$

Por lo tanto, el error aproximado es:

$$E_a = \frac{14.86249 - 15.10584}{15} = -0.01622$$

que está bastante cercano al error verdadero:

$$E_t = 14.84392 - 14.86249 = -0.01857$$

El error estimado se utiliza también para corregir la predicción

$y(2) = 14.86249 - 0.01622 = 14.84627$

la cual tiene un $E_t = -0.00235$.

25.5.2 Método de Runge-Kutta Fehlberg

Además de dividir en dos el paso, como una estrategia para ajustar el tamaño de paso, un procedimiento alternativo para obtener estimación del error consiste en calcular dos predicciones RK de diferente orden. Los resultados se restan después para obtener un estimado del error local de truncamiento. Un defecto de tal procedimiento es el gran aumento en la cantidad de cálculos. Por ejemplo, para una predicción de cuarto y quinto orden se necesita un total de 10 evaluaciones de la función por cada paso. El método de *Runge-Kutta Fehlberg* o *RK encapsulado* sagazmente evita este problema al utilizar un método RK de quinto orden que emplea las evaluaciones de la función del método RK de cuarto orden correspondiente. Así, el procedimiento genera la estimación del error ¡con sólo seis evaluaciones de la función!

En el presente caso, usamos la siguiente estimación de cuarto orden:

$$y_{i+1} = y_i + \left(\frac{37}{378}k_1 + \frac{250}{621}k_3 + \frac{125}{594}k_4 + \frac{512}{1\,771}k_6\right)h \tag{25.45}$$

junto con la fórmula de quinto orden:

$$y_{i+1} = y_i + \left(\frac{2\,825}{27\,648}k_1 + \frac{18\,575}{43\,384}k_3 + \frac{13\,525}{55\,296}k_4 + \frac{277}{14\,336}k_5 + \frac{1}{4}k_6\right)h \tag{25.46}$$

donde

$$k_1 = f(x_i, y_i)$$

$$k_2 = f\left(x_i + \frac{1}{5}h, y_i + \frac{1}{5}k_1h\right)$$

$$k_3 = f\left(x_i + \frac{3}{10}h, y_i + \frac{3}{40}k_1h + \frac{9}{40}k_2h\right)$$

$$k_4 = f\left(x_i + \frac{3}{5}h, y_i + \frac{3}{10}k_1h - \frac{9}{10}k_2h + \frac{6}{5}k_3h\right)$$

$$k_5 = f\left(x_i + h, y_i - \frac{11}{54}k_1h + \frac{5}{2}k_2h - \frac{70}{27}k_3h + \frac{35}{27}k_4h\right)$$

$$k_6 = f\left(x_i + \frac{7}{8}h, y_i + \frac{1\,631}{55\,296}k_1h + \frac{175}{512}k_2h + \frac{575}{13\,824}k_3h + \frac{44\,275}{110\,592}k_4h + \frac{253}{4\,096}k_5h\right)$$

Así, la EDO se resuelve con la ecuación (25.46) y el error estimado como la diferencia de las estimaciones de quinto y cuarto órdenes. Debemos aclarar que los coeficientes usados antes fueron desarrollados por Cash y Karp (1990). Por esta razón se le llama el método RK *Cash-Karp.*

EJEMPLO 25.13 Método de Runge-Kutta Fehlberg

Planteamiento del problema Use la versión Cash-Karp del método de Runge-Kutta Fehlberg para realizar el mismo cálculo del ejemplo 25.12 desde $x = 0$ hasta 2 usando $h = 2$.

Solución El cálculo de las *k* se resume en la siguiente tabla:

	x	*y*	*f*(*x*, *y*)
k_1	0	2	3
k_2	0.4	3.2	3.908511
k_3	0.6	4.20883	4.359883
k_4	1.2	7.228398	6.832587
k_5	2	15.42765	12.09831
k_6	1.75	12.17686	10.13237

Éstas pueden usarse para calcular la predicción de cuarto orden:

$$y_1 = 2 + \left(\frac{37}{378}3 + \frac{250}{621}4.359883 + \frac{125}{594}6.832587 + \frac{512}{1\,771}10.13237\right)2 = 14.83192$$

junto con una fórmula de quinto orden:

$$y_1 = 2 + \left(\frac{2\,825}{27\,648}3 + \frac{18\,575}{48\,384}4.359883 + \frac{13\,525}{55\,296}6.832587 + \frac{227}{14\,336}12.09831 + \frac{1}{4}10.13237\right)2 = 14.83677$$

El error estimado se obtiene al restar estas dos ecuaciones para dar:

$$E_a = 14.83677 - 14.83192 = 0.004842$$

25.5.3 Control del tamaño de paso

Ahora que hemos desarrollado formas para estimar el error de truncamiento local, se pueden usar para ajustar el tamaño de paso. En general, la estrategia es incrementar el tamaño de paso si el error es demasiado pequeño y disminuirlo si es muy grande. Press y cols. (1992) han sugerido el siguiente criterio para lograr esto:

$$h_{\text{nuevo}} = h_{\text{actual}} \left|\frac{\Delta_{\text{nuevo}}}{\Delta_{\text{actual}}}\right|^{\alpha} \tag{25.47}$$

donde h_{actual} y h_{nuevo} = tamaño de los pasos actual y nuevo, respectivamente, Δ_{actual} = exactitud actual calculada, Δ_{nuevo} = exactitud deseada, y α = exponente constante que es igual a 0.2 cuando se incrementa el tamaño de paso (por ejemplo, cuando $\Delta_{\text{actual}} \le \Delta_{\text{nuevo}}$) y a 0.25 cuando se disminuye el tamaño de paso ($\Delta_{\text{actual}} > \Delta_{\text{nuevo}}$).

El parámetro clave en la ecuación (25.47) es, obviamente, Δ_{nuevo} ya que este valor permite especificar la exactitud deseada. Una manera de lograrlo consistirá en relacionar Δ_{nuevo} con un nivel relativo de error. Como funciona bien sólo cuando se tienen valores positivos, llega a originar problemas para soluciones que pasan por cero. Por ejemplo, usted podría estar simulando una función oscilatoria que repetidamente pase por cero, pero que esté limitada por valores máximos absolutos. En tal caso, se desearía que estos valores máximos figuraran en la exactitud deseada.

Una forma más general de trabajar con estos casos es determinar Δ_{nuevo} como

$$\Delta_{\text{nuevo}} = \varepsilon y_{\text{escala}}$$

donde ε = nivel de tolerancia global. La elección de y_{escala} determinará, entonces, cómo se escala el error. Por ejemplo, si $y_{\text{escala}} = y$, la exactitud se dará en términos de errores relativos fraccionales. Si usted tiene un caso donde desee errores constantes relativos a un límite máximo preestablecido, haga y_{escala} igual a ese límite. Un truco sugerido por Press y cols. (1992) para obtener errores relativos constantes, excepto muy cerca de cero, es:

$$y_{\text{escala}} = |y| + \left| h \frac{dy}{dx} \right|$$

Ésta es la versión que usaremos en nuestro algoritmo.

25.5.4 Algoritmo computacional

Las figuras 25.21 y 25.22 muestran el pseudocódigo para implementar la versión Cash-Karp del algoritmo Runge-Kutta Fehlberg. Este algoritmo sigue el patrón dado en una implementación más detallada proporcionada por Press y cols. (1992) para sistemas de EDO.

La figura 25.21 implementa un solo paso de la rutina de Cash-Karp (que son las ecuaciones 25.45 y 25.46). La figura 25.22 muestra un programa principal general junto con una subrutina que adapta el tamaño de paso.

EJEMPLO 25.14 Aplicación con computadora de un esquema adaptativo de RK de cuarto orden

Planteamiento del problema El método adaptativo de RK es apropiado para la solución de la siguiente ecuación diferencial ordinaria

$$\frac{dy}{dx} + 0.6y = 10e^{-(x-2)^2/[2(0.075)^2]} \tag{E25.14.1}$$

Observe que para la condición inicial, $y(0) = 0.5$, la solución general es:

$$y = 0.5e^{-0.6x} \tag{E25.14.2}$$

que es una curva suave que gradualmente se aproxima a cero conforme x aumenta. En cambio, la solución particular presenta una transición abrupta en la vecindad de $x = 2$ debido a la naturaleza de la función forzada (figura 25.23*a*). Use un esquema estándar RK de cuarto orden para resolver la ecuación

```
SUBROUTINE RKkc (y,dy,x,h,yout,yerr)
PARAMETER (a2=0.2,a3=0.3,a4=0.6,a5=1.,a6=0.875,
 b21=0.2,b31=3./40.,b32=9./40.,b41=0.3,b42=–0.9,
 b43=1.2,b51=–11./54.,b52=2.5,b53=–70./27.,
 b54=35./27.,b61=1631./55296.,b62=175./512.,
 b63=575./13824.,b64=44275./110592.,b65=253./4096.,
 c1=37./378.,c3=250./621.,c4=125./594.,
 c6=512./1771.,dc1=c1–2825./27648.,
 dc3=c3–18575./48384.,dc4=c4–13525./55296.,
 dc5=–277./14336.,dc6=c6–0.25)
ytemp=y+b21*h*dy
CALL Derivs (x+a2*h,ytemp,k2)
ytemp=y+h*(b31*dy+b32*k2)
CALL Derivs(x+a3*h,ytemp,k3)
ytemp=y+h*(b41*dy+b42*k2+b43*k3)
CALL Derivs(x+a4*h,ytemp,k4)
ytemp=y+h*(b51*dy+b52*k2+b53*k3+b54*k4)
CALL Derivs(x+a5*h,ytemp,k5)
ytemp=y+h*(b61*dy+b62*k2+b63*k3+b64*k4+b65*k5)
CALL Derivs(x+a6*h,ytemp,k6)
yout=y+h*(c1*dy+c3*k3+c4*k4+c6*k6)
yerr=h*(dc1*dy+dc3*k3+dc4*k4+dc5*k5+dc6*k6)
END RKkc
```

Figura 25.21
Pseudocódigo para un solo paso del método RK Cash-Karp.

***a*) Programa principal**

```
INPUT xi, xf, yi
maxstep=100
hi=.5; tiny = 1.× 10^-30
eps=0.00005
print *, xi,yi
x=xi
y=yi
h=hi
istep=0
DO
  IF (istep > maxstep AND x ≤ xf) EXIT
  istep=istep+1
  CALL Derivs(x,y,dy)
  yscal=ABS(y)+ABS(h*dy)+tiny
  IF (x+h>xf) THEN h=xf–x
  CALL Adapt (x,y,dy,h,yscal,eps,hnxt)
  PRINT x,y
  h=hnxt
END DO
END
```

***b*) Subrutina de paso adaptativo**

```
SUB Adapt (x,y,dy,htry,yscal,eps,hnxt)
PARAMETER (safety=0.9,econ=1.89e–4)
h=htry
DO
  CALL RKkc(y,dy,x,h,ytemp,yerr)
  emax=abs(yerr/yscal/eps)
  IF emax ≤ 1 EXIT
  htemp=safety*h*emax^-0.25
  h=max(abs(htemp),0.25*abs(h))
  xnew=x+h
  IF xnew = x THEN pause
END DO
IF emax > econ THEN
  hnxt=safety*emax^-.2*h
ELSE
  hnxt=4.*h
END IF
x=x+h
y=ytemp
END Adapt
```

Figura 25.22
Pseudocódigo para: *a*) un programa principal y *b*) una subrutina de paso adaptativo para resolver una sola EDO.

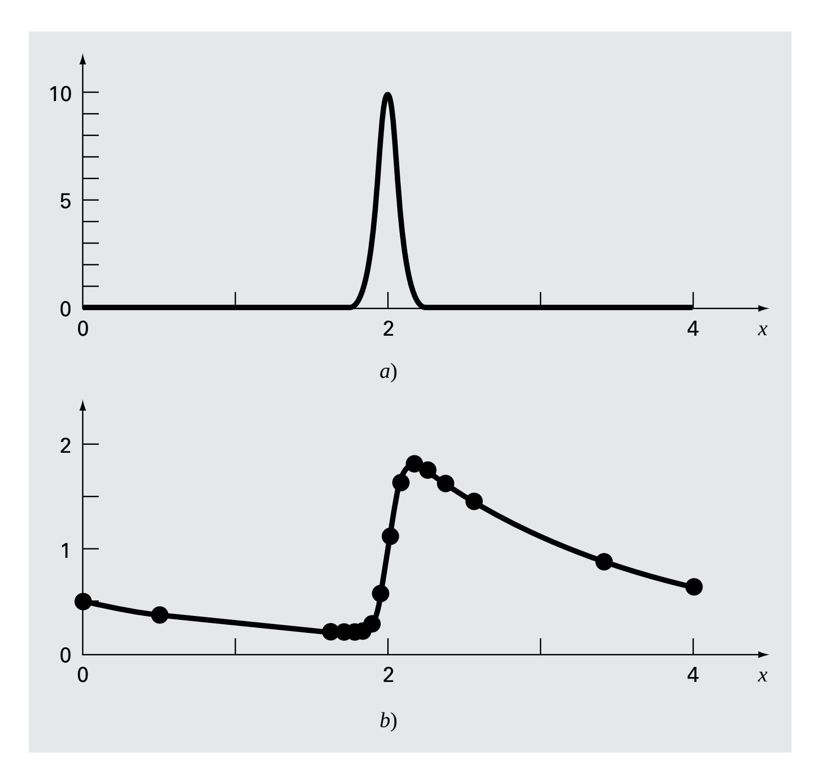


Figura 25.23
a) Una función forzada en forma de campana que induce un cambio abrupto en la solución de una EDO [ecuación (E25.14.1)]. *b*) La solución. Los puntos indican las predicciones para una rutina adaptativa paso-tamaño.

(E25.14.1) desde $x = 0$ hasta 4. Después emplee el esquema adaptativo que se describe en esta sección para realizar el mismo cálculo.

Solución Primero se utiliza el esquema clásico de cuarto orden para calcular la curva mostrada en la figura 25.23*b*. Para hacer este cálculo se usa un tamaño de paso de 0.1, de manera que se efectúan 4/(0.1) = 40 aplicaciones de la técnica. Después, se repite el cálculo con un tamaño de paso de 0.05 para un total de 80 aplicaciones. La principal discrepancia entre los dos resultados se presenta en la región que va de 1.8 a 2.0. La magnitud de la discrepancia será aproximadamente de 0.1 a 0.2 por ciento.

Después, se desarrolla el algoritmo que se muestra en las figuras 25.21 y 25.22 dentro de un programa computacional y se utiliza para resolver el mismo problema. Se elige un tamaño de paso inicial de 0.5 y una $\varepsilon = 0.00005$. Los resultados se sobreponen en la figura 25.23*b*. Observe cómo se emplean pasos grandes en las regiones de cambio gradual. Después, en la vecindad de $x = 2$, disminuyen los pasos para tomar en cuenta la naturaleza abrupta de la función forzada.

En efecto, la utilidad de un esquema de integración adaptativo depende de la naturaleza de las funciones que habrán de modelarse. En particular resulta ventajoso en aquellas soluciones con grandes tramos suaves y con regiones cortas de cambio abrupto. Además, tiene utilidad en aquellas situaciones donde no se conoce de antemano el tamaño de paso correcto. En tales casos, la rutina adaptativa "sentirá" su camino para la solución manteniendo los resultados dentro de la tolerancia deseada. Así, avanzará con pasos pequeños, "de puntillas" por regiones de cambio abrupto y acelerará el paso cuando sean más graduales las variaciones.

PROBLEMAS

25.1 Resuelva el siguiente problema de valor inicial en el intervalo de $t = 0$ a 2, donde $y(0) = 1$. Muestre todos sus resultados en la misma gráfica.

$$\frac{dy}{dt} = yt^3 - 1.1y$$

a) Analíticamente.
b) Método de Euler con $h = 0.5$ y 0.25.
c) Método de punto medio con $h = 0.5$.
d) Método RK de cuarto orden con $h = 0.5$.

25.2 Resuelva el siguiente problema en el intervalo de $x = 0$ a 1, usando un tamaño de paso de 0.25, donde $y(0) = 1$. Muestre todos sus resultados en la misma gráfica.

$$\frac{dy}{dx} = (1+4x)\sqrt{y}$$

a) Analíticamente.
b) Método de Euler.
c) Método de Heun sin iteración.
d) Método de Ralston.
e) Método RK de cuarto orden.

25.3 Utilice los métodos de *a*) Euler y *b*) Heun (sin iteración) para resolver:

$$\frac{d^2y}{dt^2} - 0.5t + y = 0$$

donde $y(0) = 2$ y $y'(0) = 0$. Resuelva de $x = 0$ a 4, con $h = 0.1$. Compare los métodos por medio de graficar las soluciones.

25.4 Resuelva el problema siguiente con el método de RK de cuarto orden:

$$\frac{d^2y}{dx^2} + 0.6\frac{dy}{dx} + 8y = 0$$

donde $y(0) = 4$ y $y'(0) = 0$. Resuelva de $x = 0$ a 5 con $h = 0.5$. Grafique sus resultados.

25.5 Resuelva de $t = 0$ a 3, con $h = 0.1$, con los métodos de *a*) Heun (sin corrector), y *b*) RK y Ralston de segundo orden:

$$\frac{dy}{dt} = y\ \text{sen}^3(t) \quad y(0) = 1$$

25.6 Solucione en forma numérica el problema siguiente, de $t = 0$ a 3:

$$\frac{dy}{dt} = -2y + t^2 \quad y(0) = 1$$

Utilice el método de RK de tercer orden, con un tamaño de paso de 0.5.

25.7 Use los métodos de *a*) Euler y *b*) RK de cuarto orden, para resolver:

$$\frac{dy}{dx} = -2y + 5e^{-t}$$

$$\frac{dz}{dx} = -\frac{yz^2}{2}$$

en el rango de $x = 0$ a 0.4, con un tamaño de paso de 0.1, con $y(0) = 2$ y $z(0) = 4$.

25.8 Calcule el primer paso del ejemplo 25.14, con el método de RK de cuarto orden adaptativo, con $h = 0.5$. Verifique si el ajuste del tamaño del paso está bien.

25.9 Si $\varepsilon = 0.001$, determine si se requiere ajustar el tamaño del paso para el ejemplo 25.12.

25.10 Use el enfoque de RK-Fehlberg para llevar a cabo el mismo cálculo del ejemplo 25.12, de $x = 0$ a 1, con $h = 1$.

25.11 Escriba un programa de computadora con base en la figura 25.7. Entre otras cosas, incluya comentarios que documenten al programa para identificar qué es lo que se pretende realizar en cada sección.

25.12 Pruebe el programa que desarrolló en el problema 25.11 para duplicar los cálculos de los ejemplos 25.1 y 25.4.

25.13 Haga un programa amigable para el usuario para el método de Heun con corrector iterativo. Pruébelo con la repetición de los resultados de la tabla 25.2.

25.14 Desarrolle un programa de computadora amigable para el usuario para el método clásico de RK de cuarto orden. Pruebe el programa con la repetición del ejemplo 25.7.

25.15 Realice un programa de computadora amigable para el usuario para sistemas de ecuaciones, con el empleo del método de RK de cuarto orden. Use este programa para duplicar el cálculo del ejemplo 25.10.

25.16 El movimiento de un sistema acoplado masa-resorte (véase la figura P25.16) está descrito por la ecuación diferencial ordinaria que sigue:

$$m\frac{d^2x}{dt^2} + c\frac{dx}{dt} + kx = 0$$

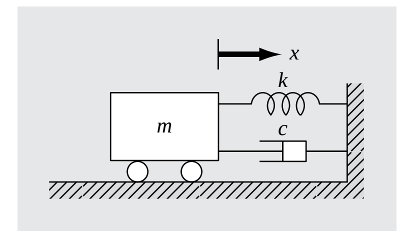


Figura P25.16

donde x = desplazamiento desde la posición de equilibrio (m), t = tiempo (s), m = 20 kg masa y c = coeficiente de amortiguamiento (N · s/m). El coeficiente de amortiguamiento c adopta tres valores, 5 (subamortiguado), 40 (amortiguamiento crítico) y 200 (sobreamortiguado). La constante del resorte es k = 20 N/m. La velocidad inicial es de cero y el desplazamiento inicial es x = 1 m. Resuelva esta ecuación con el uso de un método numérico durante el periodo $0 \le t \le 15$ s. Grafique el desplazamiento *versus* el tiempo para cada uno de los tres valores del coeficiente de amortiguamiento sobre la misma curva.

25.17 Si se drena el agua desde un tanque cilíndrico vertical por medio de abrir una válvula en la base, el líquido fluirá rápido cuando el tanque esté lleno y despacio conforme se drene. Como resultado, la tasa a la que el nivel del agua disminuye es:

$$\frac{dy}{dt} = -k\sqrt{y}$$

donde k es una constante que depende de la forma del agujero y del área de la sección transversal del tanque y agujero de drenaje. La profundidad del agua y se mide en metros y el tiempo t en minutos. Si k = 0.06, determine cuánto tiempo se requiere para vaciar el tanque si el nivel del fluido se encuentra en un inicio a 3 m. Resuelva con la aplicación de la ecuación de Euler y escriba un programa de computadora en Excel. Utilice un paso de 0.5 minutos.

25.18 La siguiente es una ecuación diferencial de segundo orden con valor inicial:

$$\frac{d^2x}{dt^2} + (5x)\frac{dx}{dt} + (x+7)\ \text{sen}(\omega t) = 0$$

donde

$$\frac{dx}{dt}(0) = 1.5 \quad y \quad x(0) = 6$$

Observe que ω = 1. Descomponga la ecuación en dos ecuaciones diferenciales de primer orden. Después de la descomposición, resuelva el sistema de t = 0 a 15, y grafique sus resultados.

25.19 Si se supone que el arrastre es proporcional al cuadrado de la velocidad, se puede modelar la velocidad de un objeto que cae, como un paracaidista, por medio de la ecuación diferencial siguiente:

$$\frac{dv}{dt} = g - \frac{c_d}{m}v^2$$

donde v es la velocidad (m/s), t = tiempo (s), g es la aceleración debida a la gravedad (9.81 m/s^2), c_d = coeficiente de arrastre de segundo orden (kg/m) y m = masa (kg). Resuelva para la velocidad y distancia que recorre un objeto de 90 kg con coeficiente de arrastre de 0.225 kg/m. Si la altura inicial es de 1 km, determine en qué momento choca con el suelo. Obtenga la solución con *a*) el método de Euler y *b*) el método de RK de cuarto orden.

25.20 Un tanque esférico tiene un orificio circular en el fondo a través del cual fluye líquido (véase la figura P25.20). La tasa de flujo a través del agujero se calcula como:

$$Q_{\text{sal}} = CA\sqrt{2gH}$$

donde Q_{sal} = flujo de salida (m^3/s), C = coeficiente obtenido en forma empírica, A = área del orificio (m^2), g = constante gravitacional (= 9.81 m/s^2) y H = profundidad del líquido dentro del tanque. Emplee alguno de los métodos numéricos descritos en este capítulo a fin de determinar cuánto tiempo tomaría que el agua fluyera por completo de un tanque de 3 m de diámetro con altura inicial de 2.75 m. Observe que el orificio tiene un diámetro de 3 cm y C = 0.55.

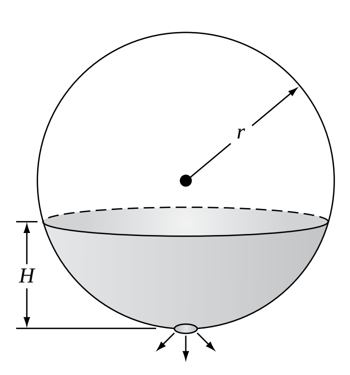


Figura P25.20
Tanque esférico.

25.21 Para simular una población se utiliza el modelo logístico:

$$\frac{dp}{dt} = k_{gm}(1 - p/p_{\text{máx}})p$$

donde p = población, k_{gm} = tasa máxima de crecimiento en condiciones ilimitadas y $p_{\text{máx}}$ es la capacidad de carga. Simule la población mundial entre 1950 y 2000, con el empleo de algún método numérico de los que se describieron en este capítulo. Para la simulación, utilice las siguientes condiciones iniciales y valores de parámetros: p_0 (en 1950) = 2 555 millones de personas, k_{gm} 0.026/año y $p_{\text{máx}}$ = 12 000 millones de personas. Haga que la función genere salidas que correspondan a las fechas de los datos siguientes de población. Desarrolle una gráfica de la simulación junto con los datos.

t	1950	1960	1970	1980	1990	2000
p	2 555	3 040	3 708	4 454	5 276	6 079

25.22 Suponga que un proyectil se lanza hacia arriba desde la superficie de la Tierra. Se acepta que la única fuerza que actúa sobre el objeto es la fuerza debida a la gravedad, hacia abajo. En estas condiciones, se usa un balance de fuerza para obtener,

$$\frac{dv}{dt} = -g(0)\frac{R^2}{(R+x)^2}$$

donde v = velocidad hacia arriba (m/s), t = tiempo (s), x = altitud (m) medida hacia arriba a partir de la superficie terrestre, $g(0)$ = aceleración gravitacional en la superficie terrestre ($\cong$ 9.81 m/s^2) y R = radio de la Tierra ($\cong 6.37 \times 10^6$ m). Como $dx/dt = v$, use el método de Euler para determinar la altura máxima que se obtendría si $v(t = 0)$ = 1 500 m/s.

25.23 La función siguiente muestra regiones tanto planas como inclinadas en una región de x relativamente corta:

$$f(x) = \frac{1}{(x-0.3)^2 + 0.01} + \frac{1}{(x-0.9)^2 + 0.04} - 6$$

Determine el valor de la integral definida de la función entre x = 0 y 1, con el método de RK adaptativo.

25.24 Dadas las condiciones iniciales, $y(0) = 1$ y $y'(0) = 0$, resuelva el problema siguiente de valor inicial de t = 0 a 4:

$$\frac{d^2y}{dt^2} + 4y = 0$$

Obtenga sus soluciones con *a*) el método de Euler y *b*) el método RK de cuarto orden. En ambos casos use tamaño de paso de 0.125. Tra-

ce ambas soluciones en la misma gráfica junto con la solución exacta $y = \cos 2t$.

25.25 Use las siguientes ecuaciones diferenciales para calcular la velocidad y posición de un balón de soccer que es pateado directo hacia arriba en el aire con una velocidad inicial de 40 m/s.

$$\frac{dy}{dt} = v$$

$$\frac{dv}{dt} = -g - \frac{c_d}{m} v|v|$$

Donde y = distancia hacia arriba (m), t = tiempo (s), v = velocidad hacia arriba (m/s), g = constante gravitacional (= 9.81 m/s^2), c_d = coeficiente de arrastre (kg/m) y m = masa (kg). Observe que el coeficiente de arrastre se relaciona a más parámetros fundamentales por medio de

$$c_d = \frac{1}{2}\rho A C_d$$

Donde ρ = densidad del aire (kg/m^3), A = área (m^2) y C_d el coeficiente de arrastre adimensional. Para los parámetros, use los siguientes valores: d = 22 cm, m = 0.4 kg, ρ = 1.3 kg/m^3 y C_d = 0.52.

25.26 En la figura P25.26 aparecen tres saltadores de bungee unidos. Si los cables de bungee están idealizados como resortes lineales (es decir, gobernados por la ley de Hooke) se pueden desarrollar las siguientes ecuaciones diferenciales basadas en los equilibrios de fuerzas.

$$m_1\frac{d^2x_1}{dt^2} = m_1g + k_2(x_2 - x_1) - k_1x_1$$

$$m_2\frac{d^2x_2}{dt^2} = m_2g + k_3(x_3 - x_2) + k_2(x_1 - x_2)$$

$$m_3\frac{d^2x_3}{dt^2} = m_3g + k_3(x_2 - x_3)$$

Donde m_i = la masa del saltador i (kg), k_j = la constante de resorte para el cable j (N/M), x_i = el desplazamiento del saltador i medido hacia abajo desde su posición de equilibrio (m) y g = aceleración gravitacional (9.81 m/s^2). Resuelva estas ecuaciones para las posiciones y velocidades de los tres saltadores dado que las condiciones iniciales de todas las posiciones y velocidades son cero en $t = 0$. Use los siguientes parámetros en sus cálculos: m_1 = 60 kg, m_2 = 70 kg, m_3 = 80 kg, $k_1 = k_3$ = 50 y k_2 = 100 (N/m).

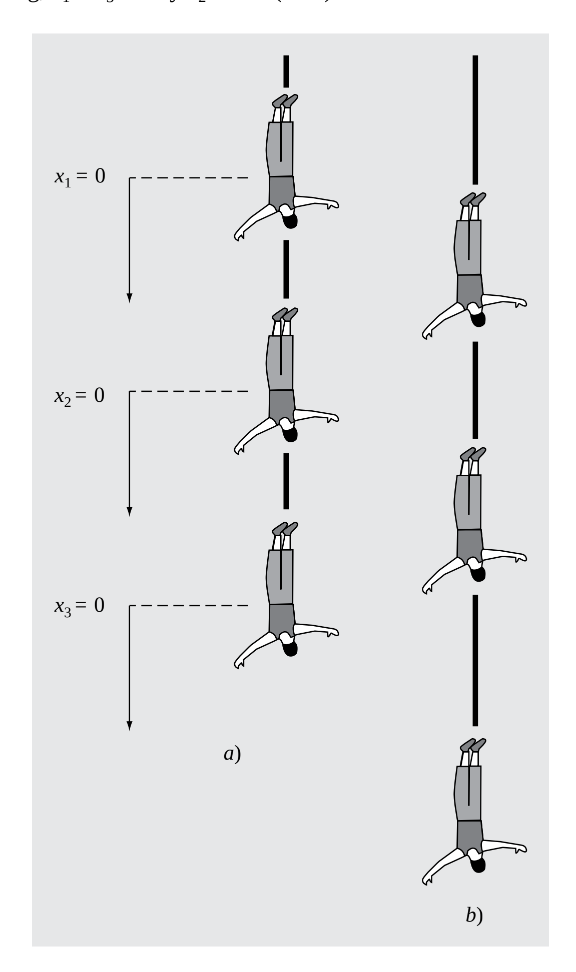


Figura P25.26
Tres individuos unidos con cables de bungee. *a*) No estirado, *b*) Estirado.

CAPÍTULO 26

Métodos rígidos y de pasos múltiples

El presente capítulo cubre dos áreas de estudio. Primero, describimos las EDO *rígidas*. Éstas son tanto EDO en forma individuales como sistemas de EDO, que tienen componentes rápidos y lentos para su solución. Presentamos la idea de una técnica de *solución implícita* como una respuesta comúnmente utilizada para este problema. Después analizamos los *métodos de pasos múltiples* o *multipaso*. Estos algoritmos guardan información de pasos anteriores para obtener de manera más efectiva la trayectoria de la solución; también ofrecen la estimación del error de truncamiento que se utiliza para implementar el control adaptativo del tamaño de paso.

26.1 RIGIDEZ

El término rigidez constituye un problema especial que puede surgir en la solución de ecuaciones diferenciales ordinarias. Un *sistema rígido* es aquel que tiene componentes que cambian rápidamente, junto con componentes de cambio lento. En muchos casos, los componentes de variación rápida son efímeros, transitorios, que desaparecen, después de lo cual la solución es dominada por componentes de variación lenta. Aunque los fenómenos transitorios existen sólo en una pequeña parte del intervalo de integración, pueden determinar el tiempo en toda la solución.

Tanto las EDO individuales como los sistemas pueden ser rígidos. Un ejemplo de una EDO rígida es:

$$\frac{dy}{dt} = -1\,000y + 3\,000 - 2\,000e^{-t} \tag{26.1}$$

Si $y(0) = 0$, la solución analítica que se obtiene es:

$$y = 3 - 0.998e^{-1\,000t} - 2.002e^{-t} \tag{26.2}$$

Como se muestra en la figura 26.1, la solución al principio se encuentra dominada por el término exponencial rápido ($e^{-1\,000t}$). Después de un periodo muy corto ($t < 0.005$), esta parte transistoria termina y la solución se regirá por el exponencial lento (e^{-t}).

Al examinar la parte homogénea de la ecuación (26.l), se conoce el tamaño de paso necesario para la estabilidad de tal solución:

$$\frac{dy}{dt} = -ay \tag{26.3}$$

Si $y(0) = y_0$, puede usarse el cálculo para determinar la solución

$$y = y_0e^{-at}$$

Así, la solución empieza en y_0 y asintóticamente se aproxima a cero.

Es factible usar el método de Euler para resolver el mismo problema en forma numérica:

$$y_{i+1} = y_i + \frac{dy_i}{dt}h$$

Al sustituir la ecuación (26.3) se tiene

$$y_{i+1} = y_i - ay_ih$$

o

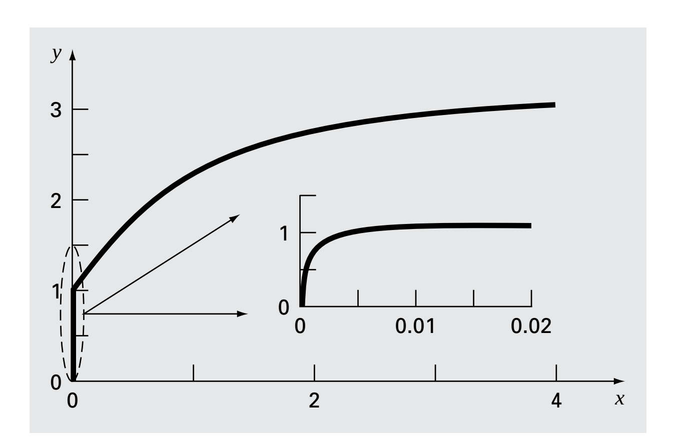


Figura 26.1
Gráfica de una solución rígida para una sola EDO. Aunque la solución parece iniciar en 1, en realidad existe una forma transitoria rápida desde $y = 0$ hasta 1, que ocurre en menos de 0.005 unidades de tiempo. Esta transición es perceptible sólo cuando la respuesta se observa sobre una escala de tiempo más fina en ese intervalo.

$$y_{i+1} = y_i(1 - ah) \tag{26.4}$$

La estabilidad de esta fórmula, sin duda, depende del tamaño de paso h. Es decir, $|1 - ah|$ debe ser menor que 1. Entonces, si $h > 2/a$, $|y_i| \to \infty$ conforme $i \to \infty$.

En la parte transitoria rápida de la ecuación (26.2) se utiliza este criterio con la finalidad de mostrar que para mantener la estabilidad el tamaño de paso debe ser < 2/1 000 = 0.002. Además, deberá observarse que mientras este criterio mantiene la estabilidad (es decir, una solución acotada), sería necesario un tamaño de paso aún más pequeño para obtener una solución exacta. Así, aunque la parte transitoria se presenta sólo en una pequeña fracción del intervalo de integración, ésta controla el tamaño de paso máximo permitido.

Sin ahondar mucho, se podrá suponer que las rutinas adaptativas de tamaño de paso descritas al final del capítulo ofrecerán una solución a este problema. Quizá pensará que tales rutinas usarían pasos pequeños en las partes transistorias rápidas y pasos grandes en las otras. Sin embargo, éste no es el caso, ya que para los requerimientos de estabilidad se necesitarán pasos muy pequeños en toda la solución.

En lugar de usar procedimientos explícitos, los métodos implícitos ofrecen una solución alternativa. Tales representaciones se denominan *implícitas*, debido a que la incógnita aparece en ambos lados de la ecuación. Una forma implícita del método de Euler se desarrolla evaluando la derivada en el tiempo futuro,

$$y_{i+1} = y_i + \frac{dy_{i+1}}{dt}h$$

A esto se le llama: *método de Euler hacia atrás* o *implícito*. Si se sustituye la ecuación (26.3) se llega a:

$$y_{i+1} = y_i - ay_{i+1}h$$

de donde se obtiene:

$$y_{i+1} = \frac{y_i}{1 + ah} \tag{26.5}$$

En este caso, sin importar el tamaño de paso, $|y_i| \to 0$ conforme $i \to \infty$. De ahí que el procedimiento se llame *incondicionalmente estable*.

EJEMPLO 26.1 Euler explícito e implícito

Planteamiento del problema Con los métodos explícito e implícito de Euler resuelva

$$\frac{dy}{dt} = -1\,000y + 3\,000 - 2\,000e^{-t}$$

donde $y(0) = 0$. *a*) Use el método de Euler explícito con tamaños de paso de 0.0005 y 0.0015 para encontrar y entre $t = 0$ y 0.006. *b*) Utilice el método implícito de Euler con un tamaño de paso de 0.05 para encontrar y entre 0 y 0.4.

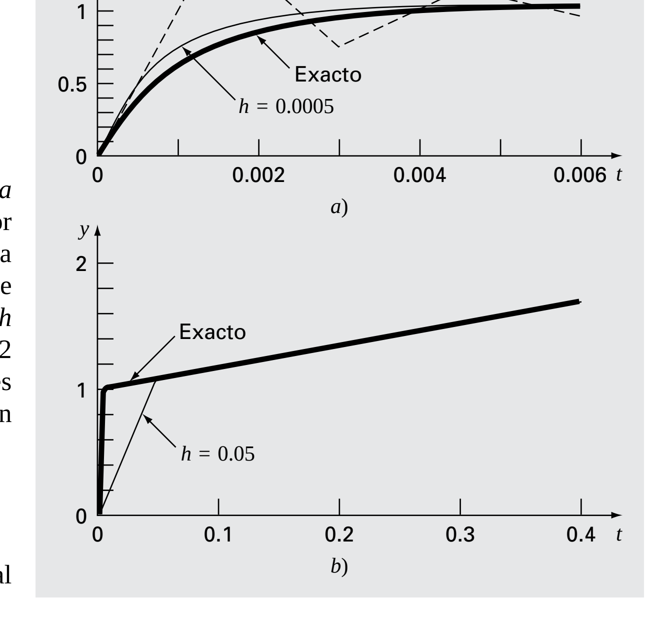


Figura 26.2
Solución de una EDO "rígida" con los métodos de Euler *a*) explícito y *b*) implícito.

Solución

a) En este problema, el método explícito de Euler es:

$$y_{i+1} = y_i + (-1\,000y_i + 3\,000 - 2\,000e^{-t_i})h$$

El resultado para $h = 0.0005$ se despliega en la figura 26.2*a* junto con la solución analítica. Aunque muestre algún error de truncamiento, el resultado capta la forma general de la solución analítica. En cambio, cuando el tamaño de paso se incrementa a un valor justo debajo del límite de estabilidad ($h = 0.0015$), la solución presenta oscilaciones. Usando $h > 0.002$ se tiene como resultado una solución totalmente inestable; es decir, la solución tenderá al infinito conforme se avanza en las iteraciones.

b) El método de Euler implícito es:

$$y_{i+1} = y_i + (-1\,000y_{i+1} + 3\,000 - 2\,000e^{-t_{i+1}})h$$

Ahora como la EDO es lineal, se reordena esta ecuación de tal forma que y_{i+1} quede sola en el lado izquierdo,

$$y_{i+1} = \frac{y_i + 3\,000h - 2\,000he^{-t_{i+1}}}{1 + 1\,000h}$$

El resultado con $h = 0.05$ se muestra en la figura 26.2*b* junto con la solución analítica. Observe que aun cuando usamos un tamaño de paso mucho mayor que aquel que indujo la inestabilidad en el método de Euler explícito, la solución numérica se ajusta muy bien al resultado analítico.

Los sistemas de EDO también pueden ser rígidos. Un ejemplo es:

$$\frac{dy_1}{dt} = -5y_1 + 3y_2 \tag{26.6a}$$

$$\frac{dy_2}{dt} = 100y_1 - 301y_2 \tag{26.6b}$$

Para las condiciones iniciales $y_1(0) = 52.29$ y $y_2(0) = 83.82$, la solución exacta es:

$$y_1 = 52.96e^{-3.9899t} - 0.67e^{-302.0101t} \tag{26.7a}$$

$$y_2 = 17.83e^{-3.9899t} + 65.99e^{-302.0101t} \tag{26.7b}$$

Observe que los exponentes son negativos y difieren por cerca de 2 órdenes de magnitud. Como en una sola ecuación, los exponentes grandes son los que responden rápidamente y representan la esencia de la rigidez del sistema.

Para este ejemplo el método implícito de Euler para sistemas se formula como

$$y_{1,i+1} = y_{1,i} + (-5y_{1,i+1} + 3y_{2,i+1})h \tag{26.8a}$$

$$y_{2,i+1} = y_{2,i} + (100y_{1,i+1} - 301y_{2,i+1})h \tag{26.8b}$$

Al agrupar términos se tiene

$$(1 + 5h)y_{1,i+1} - 3hy_{2,i+1} = y_{1,i} \tag{26.9a}$$

$$-100hy_{1,i+1} + (1 + 301h)y_{2,i+1} = y_{2,i} \tag{26.9b}$$

Así, notamos que el problema consiste en resolver un conjunto de ecuaciones simultáneas en cada paso.

Para EDO no lineales, la solución se vuelve aún más difícil, ya que debe resolverse un sistema de ecuaciones simultáneas no lineales (recuerde la sección 6.6). Así, aunque se gana estabilidad a través de procedimientos implícitos, se paga un precio al agregar mayor complejidad a la solución.

El método implícito de Euler es incondicionalmente estable y tiene sólo una exactitud de primer orden. También es posible desarrollar de manera similar un esquema de integración para la regla del trapecio implícita de segundo orden para sistemas rígidos. En general, es preferible tener métodos de orden superior. Las fórmulas de Adams-Moulton descritas después, en este capítulo, también son útiles para determinar métodos implícitos de orden superior. Sin embargo, los límites de estabilidad de tales procedimientos son muy rigurosos cuando se aplican a sistemas rígidos. Gear (1971) desarrolló una serie especial de esquemas implícitos que tienen límites de estabilidad más grandes, basados en las fórmulas de diferencias hacia atrás. Se ha hecho un trabajo muy fuerte para desarrollar el software que implemente los métodos de Gear en forma eficiente. Dando como resultado que sea probablemente el método más utilizado para resolver sistemas rígidos. Además, Rosenbrock y otros (véase Press y colaboradores, 2007) han propuesto algoritmos implícitos, de Runge-Kutta, donde los términos k aparecen en forma implícita. Dichos métodos poseen buenas características de estabilidad y son bastante adecuados para resolver sistemas de ecuaciones diferenciales ordinarias rígidas.

26.2 MÉTODOS DE PASOS MÚLTIPLES

Los métodos de un paso que se describieron en las secciones anteriores utilizan información de un solo punto x_i para predecir un valor de la variable dependiente, y_{i+1}, en un valor futuro, de la variable x_{i+1} (figura 26.3*a*). Los procedimientos alternativos, llamados *métodos de pasos múltiples* o *multipasos* (figura 26.3*b*), se basan en que, una vez empezado el cálculo, se tiene a disposición información de los puntos anteriores. La curvatura de las líneas que unen esos valores previos ofrecen información respecto a la trayectoria de la solución. Los métodos de pasos múltiples explorados en este capítulo aprovechan tal información para resolver las EDO. Antes de describir las versiones de orden superior, presentaremos un método simple de segundo orden que sirve para demostrar las características generales de los procedimientos multipaso.

26.2.1 El método de Heun sin autoinicio

Recuerde que el procedimiento de Heun utiliza el *método de Euler* como un *predictor* [ecuación (25.15)]:

$$y_{i+1}^0 = y_i + f(x_i, y_i)h \tag{26.10}$$

y la *regla del trapecio* como un *corrector* [ecuación (25.16)]:

$$y_{i+1} = y_i + \frac{f(x_i, y_i) + f(x_{i+1}, y_{i+1}^0)}{2}h \tag{26.11}$$

Así, el predictor y el corrector tienen errores de truncamiento local de $O(h^2)$ y $O(h^3)$, respectivamente. Esto sugiere que el predictor es la parte débil en el método, a causa de que tiene el error más grande. Esta debilidad es significativa puesto que la eficiencia del paso corrector depende de la exactitud de la predicción inicial. En consecuencia, una

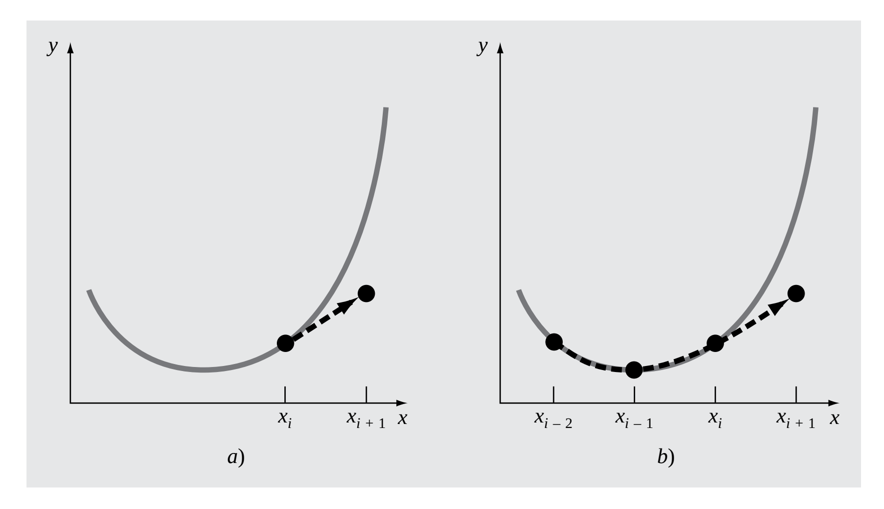


Figura 26.3
Representación gráfica de la diferencia fundamental entre los métodos *a*) de un paso y *b*) de pasos múltiples para la solución de EDO.

forma de mejorar el método de Heun consiste en desarrollar un predictor que tenga un error local de $O(h^3)$. Esto se obtiene usando el método de Euler y la pendiente en y_i, así como información extra de un valor anterior y_{i-1} como en:

$$y_{i+1}^0 = y_{i-1} + f(x_i, y_i)2h \tag{26.12}$$

Observe que la ecuación (26.12) alcanza $O(h^3)$ a expensas de emplear un tamaño de paso mayor, $2h$. Además, note que la ecuación (26.12) no es de autoinicio, ya que necesita un valor previo de la variable dependiente y_{i-1}. Tal valor no está disponible en un problema común de valor inicial. Por ello, las ecuaciones (26.11) y (26.12) se denominan *método de Heun sin autoinicio.*

Como se indica en la figura 26.4, la derivada estimada para la ecuación (26.12) se localiza ahora en el punto medio y no al inicio del intervalo sobre el cual se hace predicción. Como se consideró anteriormente, esta ubicación centrada mejora el error del predictor a $O(h^3)$. Sin embargo, antes de realizar una deducción formal del método de Heun sin autoinicio, lo resumiremos y lo expresaremos utilizando una nomenclatura ligeramente modificada:

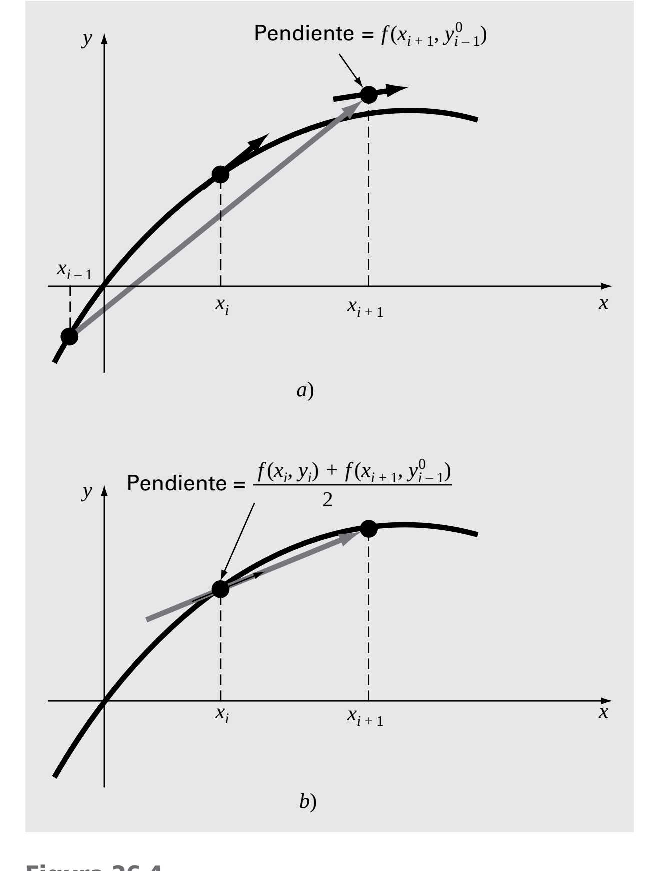


Figura 26.4
Representación gráfica del método de Heun sin autoinicio. *a*) El método de punto medio que se utiliza como un predictor. *b*) La regla del trapecio que se emplea como un corrector.

Predictor: $$y_{i+1}^0 = y_{i-1}^m + f(x_i, y_i^m)2h \tag{26.13}$$

Corrector: $$y_{i+1}^j = y_i^m + \frac{f(x_i, y_i^m) + f(x_{i+1}, y_{i+1}^{j-1})}{2}h \tag{26.14}$$

(para $j = 1, 2, \ldots, m$)

donde los superíndices se agregaron para denotar que el corrector se aplica iterativamente desde $j = 1$ hasta m para obtener mejores soluciones. Observe que y_i^m y y_{i-1}^m son los resultados finales del corrector en los pasos anteriores. Las iteraciones terminan en cualquier paso considerando el criterio de terminación:

$$|\varepsilon_a| = \left| \frac{y_{i+1}^j - y_{i+1}^{j-1}}{y_{i+1}^j} \right| 100\% \tag{26.15}$$

Cuando ε_a es menor que una tolerancia de error ε_s preestablecida, concluyen las iteraciones. En este momento, $j = m$. El uso de las ecuaciones (26.13) a (26.15) para resolver una EDO se demuestra en el siguiente ejemplo.

EJEMPLO 26.2 Método de Heun sin autoinicio

Planteamiento del problema Con el método de Heun sin autoinicio realice los mismos cálculos como en el ejemplo 25.5 donde se usó el método de Heun. Es decir, integre $y' = 4e^{0.8x} - 0.5y$ y desde $x = 0$ hasta $x = 4$ con un tamaño de paso de 1.0. Igual que en el ejemplo 25.5, la condición inicial en $x = 0$ es $y = 2$. Sin embargo, como aquí tenemos un método de pasos múltiples, requerimos de información adicional, considerando que $y = -0.3929953$ en $x = -1$.

Solución El predictor [ecuación (26.13)] se utiliza para extrapolar linealmente de $x = -1$ a $x = 1$.

$$y_1^0 = -0.3929953 + [4e^{0.8(0)} - 0.5(2)]2 = 5.607005$$

El corrector [ecuación (26.14)] se usa después para calcular el valor:

$$y_1^1 = 2 + \frac{4e^{0.8(0)} - 0.5(2) + 4e^{0.8(1)} - 0.5(5.607005)}{2}1 = 6.549331$$

que representa un error relativo porcentual de –5.73% (valor verdadero = 6.194631). Este error es más pequeño que el valor de –8.18% en el que se incurre con el método de Heun de autoinicio.

Ahora, se aplica la ecuación (26.14) de manera iterativa para mejorar la solución:

$$y_1^2 = 2 + \frac{3 + 4e^{0.8(1)} - 0.5(6.549331)}{2}1 = 6.313749$$

que representa un ε_t de –1.92%. Se determina un estimado del error utilizando la ecuación (26.15):

$$|\varepsilon_a| = \left|\frac{6.313749 - 6.549331}{6.313749}\right| 100\% = 3.7\%$$

La ecuación (26.14) se aplica de manera iterativa hasta que ε_a esté por debajo de un valor preespecificado de ε_s. Como fue el caso con el método de Heun (recuerde el ejemplo 25.5), las iteraciones convergen a un valor de 6.360865 (ε_t = –2.68%). Sin embargo, como el valor del predictor inicial es más exacto, el método de pasos múltiples converge más rápido.

En el segundo paso, el predictor es:

$$y_2^0 = 2 + [4e^{0.8(1)} - 0.5(6.360865)]2 = 13.44346 \qquad \varepsilon_t = 9.43\%$$

el cual es mejor que la predicción de 12.08260 (ε_t = 18%) calculada con el método de Heun original. El primer corrector da 15.76693 (ε_t = 6.8%); las siguientes iteraciones convergen al mismo resultado como en el método de Heun de autoinicio: 15.30224 (ε_t = –3.1%). Observe que en el paso anterior, la rapidez de convergencia del corrector es mayor debido a la mejoría de la predicción inicial.

Deducción y análisis del error de las fórmulas del predictor-corrector Ya empleamos conceptos gráficos para deducir el método de Heun sin autoinicio. Ahora mostraremos cómo las mismas ecuaciones se pueden deducir en forma matemática. Tal deducción es interesante en especial porque vincula los conceptos de ajuste de curvas, de integración numérica y de EDO. La deducción también es útil porque ofrece un procedimiento simple para desarrollar métodos de pasos múltiples de orden superior y una estimación de sus errores.

La deducción se basa en resolver la EDO general

$$\frac{dy}{dx} = f(x, y)$$

Esta ecuación se resuelve multiplicando ambos lados por dx e integrando entre los límites i e i + 1:

$$\int_{y_i}^{y_{i+1}} dy = \int_{x_i}^{x_{i+1}} f(x, y)\, dx$$

El lado izquierdo se integra y evalúa mediante [recuerde la ecuación (25.21)]:

$$y_{i+1} = y_i + \int_{x_i}^{x_{i+1}} f(x, y)\, dx \tag{26.16}$$

La ecuación (26.16) representa una solución a la EDO si la integral puede evaluarse. Es decir, proporciona un medio para calcular un nuevo valor de la variable dependiente y_{i+1} a partir de un valor previo y_i y del conocimiento de la ecuación diferencial.

Las fórmulas de integración numérica como las que se desarrollaron en el capítulo 21 proporcionan una manera de realizar esta evaluación. Por ejemplo, la regla del trapecio [ecuación (21.3)] se utiliza para evaluar la integral, como sigue:

$$\int_{x_i}^{x_{i+1}} f(x, y)\, dx = \frac{f(x_i, y_i) + f(x_{i+1}, y_{i+1})}{2} h \tag{26.17}$$

donde $h = x_{i+1} - x_i$ es el tamaño de paso. Sustituyendo la ecuación (26.17) en la ecuación (26.16) se tiene:

$$y_{i+1} = y_i + \frac{f(x_i, y_i) + f(x_{i+1}, y_{i+1})}{2} h$$

que es el paso corrector para el método de Heun. Como esta ecuación se basa en la regla del trapecio, el error de truncamiento se puede tomar directamente de la tabla 21.2:

$$E_c = -\frac{1}{12}h^3 y^{(3)}(\xi_c) = -\frac{1}{12}h^3 f''(\xi_c) \tag{26.18}$$

donde el subíndice c indica que éste es el error del corrector.

Se utiliza un procedimiento similar para obtener el predictor. En este caso, los límites de integración serán $i - 1$ e $i + 1$:

$$\int_{y_{i-1}}^{y_{i+1}} dy = \int_{x_{i-1}}^{x_{i+1}} f(x, y)\, dx$$

que se integra y se reordena para tener

$$y_{i+1} = y_{i-1} + \int_{x_{i-1}}^{x_{i+1}} f(x,y)\, dx \tag{26.19}$$

Ahora, no se emplea una fórmula cerrada de la tabla 21.2, sino que se utiliza la primera fórmula de integración abierta de Newton-Cotes (véase tabla 21.4) para evaluar la integral, como sigue:

$$\int_{x_{i-1}}^{x_{i+1}} f(x,y)\,dx = 2hf(x_i, y_i) \tag{26.20}$$

que se llama *método del punto medio.* Sustituyendo la ecuación (26.20) en la ecuación (26.19) se obtiene:

$$y_{i+1} = y_{i-1} + 2hf(x_i, y_i)$$

el cual es el predictor para el método de Heun sin autoinicio. Como en el corrector, el error de truncamiento local se puede tomar directamente de la tabla 21.4:

$$E_p = \frac{1}{3}h^3 y^{(3)}(\xi_p) = \frac{1}{3}h^3 f''(\xi_p) \tag{26.21}$$

donde el subíndice p indica que éste es el error del predictor.

Así, el predictor y el corrector en el método de Heun sin autoinicio tiene errores de truncamiento del mismo orden. Además de actualizar la exactitud del predictor, este hecho tiene ventajas adicionales relacionadas con el análisis del error, como se verá en la siguiente sección.

Estimación de errores Si el predictor y el corrector de un método de pasos múltiples son del mismo orden, el error de truncamiento local puede estimarse en el proceso de cada cálculo. Esto representa una enorme ventaja, ya que establece un criterio para el ajuste del tamaño de paso.

El error de truncamiento local del predictor se estima con la ecuación (26.21). Dicho error estimado se combina con el estimado de y_{i+1} del paso predictor para dar [recuerde nuestra definición básica en la ecuación (3.1)]:

$$\text{Valor verdadero} = y_{i+1}^0 + \frac{1}{3}h^3 y^{(3)}(\xi_p) \tag{26.22}$$

Usando un procedimiento similar, el error estimado para el corrector [ecuación (26.18)] se combina con el resultado del corrector y_{i+1} para llegar a:

$$\text{Valor verdadero} = y_{i+1}^m - \frac{1}{12}h^3 y^{(3)}(\xi_c) \tag{26.23}$$

La ecuación (26.22) se resta de la ecuación (26.23) para tener:

$$0 = y_{i+1}^m - y_{i+1}^0 - \frac{5}{12}h^3 y^{(3)}(\xi) \tag{26.24}$$

donde ξ está ahora entre x_{i-1} y x_{i+1}. Ahora, si se divide la ecuación (26.24) entre 5 y se reordena el resultado se obtiene:

$$\frac{y_{i+1}^0 - y_{i+1}^m}{5} = -\frac{1}{12} h^3 y^{(3)}(\xi) \tag{26.25}$$

Observe que los lados derechos de las ecuaciones (26.18) y (26.25) son idénticos, con excepción del argumento de la tercera derivada. Si la tercera derivada no tiene una variación apreciable en el intervalo dado, supondremos que los lados derechos son iguales y, por lo tanto, los lados izquierdos deberían ser equivalentes, como:

$$E_c = -\frac{y_{i+1}^0 - y_{i+1}^m}{5} \tag{26.26}$$

Así, llegamos a una relación que puede utilizarse para estimar el error de truncamiento por paso con base en dos cantidades [el predictor (y_{i+1}^0) y el corrector (y_{i+1}^m)], que son producidos durante el cálculo.

EJEMPLO 26.3 Estimación del error de truncamiento por paso

Planteamiento del problema Con la ecuación (26.26) estime el error de truncamiento por paso del ejemplo 26.2. Observe que los valores verdaderos en $x = 1$ y 2 son 6.194631 y 14.84392, respectivamente.

Solución En $x_{i+1} = 1$, el predictor es 5.607005 y el corrector es 6.360865. Se sustituyen estos valores en la ecuación (26.26):

$$E_c = -\frac{6.360865 - 5.607005}{5} = -0.1507722$$

que no está lejos del error exacto,

$$E_t = 6.194631 - 6.360865 = -0.1662341$$

En $x_{i+1} = 2$, el predictor es 13.44346 y el corrector es 15.30224, que se utiliza para calcular:

$$E_c = -\frac{15.30224 - 13.44346}{5} = -0.3717550$$

que tampoco está lejos del error exacto, $E_t = 14.84392 - 15.30224 = -0.4583148$.

La facilidad con que se estima el error mediante la ecuación (26.26) proporciona una buena forma de ajustar el tamaño de paso, durante el proceso de cada cálculo. Por ejemplo, si la ecuación (26.26) indica que el error es mayor que un nivel aceptable, el tamaño de paso podrá disminuirse.

Modificadores Antes de analizar los algoritmos de cómputo, es necesario observar otras dos maneras en que el método de Heun sin autoinicio puede volverse más exacto y eficiente. Primero, habrá que percatarse de que además de ofrecer un criterio para el ajuste del tamaño de paso, la ecuación (26.26) representa una estimación numérica de la discrepancia entre el valor final corregido en cada paso y_{i+1} y el valor verdadero. Así, ésta puede sumarse directamente a y_{i+1} para mejorar aún más el estimado:

$$y_{i+1}^m \leftarrow y_{i+1}^m - \frac{y_{i+1}^m - y_{i+1}^0}{5} \tag{26.27}$$

La ecuación (26.27) se conoce como *modificador del corrector*. (El símbolo ← se lee "es reemplazado por".) El lado izquierdo es el valor modificado de y_{i+1}^m.

Una segunda mejoría relacionada más con la eficiencia del programa es un *modificador del predictor*, que está diseñado para ajustar el resultado del predictor de forma que está más cerca al valor convergente final del corrector. Esto resulta ventajoso debido a que, como se observó al inicio de esta sección, el número de iteraciones del corrector es altamente dependiente de la exactitud de

la predicción inicial. En consecuencia, si la predicción se modifica en forma adecuada, podríamos reducir el número de iteraciones necesarias para converger al último valor del corrector.

Tal modificador puede deducirse en forma sencilla al suponer que la tercera derivada es relativamente constante de un paso a otro. Por lo tanto, usando el resultado del paso previo en i, de la ecuación (26.25) se puede despejar

$$h^3 y^{(3)}(\xi) = -\frac{12}{5}(y_i^0 - y_i^m) \tag{26.28}$$

suponiendo que $y^{(3)}(\xi) \cong y^{(3)}(\xi_p)$, se sustituye en la ecuación (26.21) para dar

$$E_p = \frac{4}{5}(y_i^m - y_i^0) \tag{26.29}$$

que se utiliza después para modificar el resultado del predictor:

$$y_{i+1}^0 \leftarrow y_{i+1}^0 + \frac{4}{5}(y_i^m - y_i^0) \tag{26.30}$$

EJEMPLO 26.4 Efecto de los modificadores sobre los resultados del predictor-corrector

Planteamiento del problema Vuelva a calcular el ejemplo 26.3 empleando ambos modificadores.

Solución Como en el ejemplo 26.3, el resultado del predictor inicial es 5.607005. Ya que el modificador del predictor [ecuación (26.30)] requiere valores de una iteración previa, no es posible emplearlo para mejorar este resultado inicial. Sin embargo, la ecuación (26.27) sirve para modificar el valor corregido de 6.360865 (ε_t = –2.684%), como sigue:

$$y_1^m = 6.360865 - \frac{6.360865 - 5.607005}{5} = 6.210093$$

que representa un ε_t = –0.25%. Así, el error se reduce en un orden de magnitud.

En la siguiente iteración, el predictor [ecuación (26.13)] se usa para calcular

$$y_2^0 = 2 + [4e^{0.8(0)} - 0.5(6.210093)]2 = 13.59423 \qquad \varepsilon_t = 8.42\%$$

que es aproximadamente la mitad del error del predictor para la segunda iteración del ejemplo 26.3, es decir, ε_t = 18.6%. Esta mejoría ocurre debido a que utilizamos aquí una mejor estimación de y (6.210093 en lugar de 6.360865) en el predictor. En otras palabras, los errores propagado y global se reducen al incluir el modificador del corrector.

Ahora debido a que tenemos información de la iteración anterior, la ecuación (26.30) se emplea para modificar el predictor, como sigue:

$$y_2^0 = 13.59423 + \frac{4}{5}(6.360865 - 5.607005) = 14.19732 \qquad \varepsilon_t = -4.36\%$$

que, de nuevo, reduce el error a la mitad.

Esta modificación no tiene efecto en el resultado final del siguiente paso del corrector. Sin importar si se usan los predictores modificados o no modificados, el corrector, al final, convergerá a la misma respuesta. No obstante, como la rapidez o eficiencia de convergencia depende de la exactitud de la predicción inicial, la modificación puede reducir el número de iteraciones requerido para la convergencia.

La implementación del corrector da un resultado de 15.21178 (ε_t = –2.48%), el cual representa una mejora sobre el ejemplo 26.3 debido a la reducción del error global. Por último, tal resultado se puede modificar usando la ecuación (26.27):

$$y_2^m = 15.21178 - \frac{15.21178 - 13.59423}{5} = 14.88827 \qquad \varepsilon_t = -0.30\%$$

De nuevo, el error se redujo en un orden de magnitud.

Como en el ejemplo anterior, al incluir los modificadores se incrementó tanto la eficiencia como la exactitud de los métodos de pasos múltiples. En particular, el modificador del corrector incrementa efectivamente el orden de la técnica. Así, el método de Heun sin autoinicio con modificadores es de tercer orden y no de segundo orden, como en el caso de la versión no modificada. Aunque deberá observarse que hay situaciones donde el modificador del corrector afectará la estabilidad del proceso de iteración del corrector. En consecuencia, el modificador no se incluye en el algoritmo de Heun sin autoinicio que se presenta en la figura 26.5. A menos que el modificador del corrector todavía pueda tener utilidad en el control del tamaño de paso, como se analizará después.

26.2.2 Control del tamaño de paso y programas computacionales

Tamaño de paso constante Es relativamente simple desarrollar una versión con tamaño de paso constante del método de Heun sin autoinicio. La única complicación es que se requiere de un método de un paso para generar el punto extra necesario para iniciar el cálculo.

Además, como se emplea un tamaño de paso constante, se debe elegir un valor para h antes de los cálculos. En general, la experiencia indica que un tamaño de paso óptimo deberá ser lo suficientemente pequeño para asegurar la convergencia con dos iteraciones del corrector (Hull y Creemer, 1963). Además, debe ser lo suficientemente pequeño para dar un error de truncamiento lo suficientemente pequeño. Al mismo tiempo, el tamaño de paso deberá ser tan grande como sea posible para minimizar el costo de ejecución y el error de redondeo. Como se hizo con los otros métodos para EDO, la única forma práctica para evaluar la magnitud del error global es comparar los resultados del mismo problema utilizando la mitad del tamaño de paso.

Tamaño de paso variable Normalmente se utilizan dos criterios para decidir si se justifica un cambio de tamaño de paso. Primero, si la ecuación (26.26) es mayor que algún criterio de error preespecificado, se disminuye el tamaño de paso. Segundo, se elige el tamaño de paso de manera tal que el criterio de convergencia del corrector se satisfaga con dos iteraciones. Este criterio busca tomar en cuenta las ventajas y las desventajas entre la rapidez de convergencia y el número total de pasos en el cálculo. Para valores más pequeños de h la convergencia será más rápida, pero se requieren más pasos. Para valores más grandes de h la convergencia se vuelve más lenta, pero se tiene un menor número de pasos. La experiencia (Hull y Creemer, 1963) sugiere que el total de pasos se minimizará si h se elige de tal forma que el corrector converja en dos iteraciones. Por lo

Predictor:

$$y_{i+1}^0 = y_{i-1}^m + f(x_i, y_i^m)2h$$

(Guarde el resultado como $y_{i+1,u}^0 = y_{i+1}^0$, donde el subíndice u designa que la variable no está modificada.)

Modificador del predictor:

$$y_{i+1}^0 \leftarrow y_{i+1,u}^0 + \frac{4}{5}(y_{i,u}^m - y_{i,u}^0)$$

Corrector:

$$y_{i+1}^j = y_i^m + \frac{f(x_i, y_i^m) + f(x_{i+1}, y_{i+1}^{j-1})}{2}h \qquad \text{(para } j = 1 \text{ a máximo } m \text{ iteraciones)}$$

Verificación del error:

$$|\varepsilon_a| = \left|\frac{y_{i+1}^j - y_{i+1}^{j-1}}{y_{i+1}^j}\right| 100\%$$

(Si $|\varepsilon_a|$ > criterio de error, hacer $j = j + 1$ y repita el corrector; si $\varepsilon_a \le$ criterio de error, guarde el resultado como $y_{i+1,u}^m = y_{i+1}^m$.)

Estimación del error del corrector:

$$E_c = -\frac{1}{5}(y_{i+1,u}^m - y_{i+1,u}^0)$$

(Si el cálculo continúa, hacer $i = i + 1$ y regrese al predictor.)

Figura 26.5
Secuencia de fórmulas usadas para implementar el método de Heun sin autoinicio. Observe que es posible utilizar la estimación del error del corrector para modificar el corrector. Sin embargo, como esto llega a afectar la estabilidad del corrector, el modificador no se incluye en este algoritmo. La estimación del error del corrector se incluye debido a su utilidad en el ajuste del tamaño de paso.

tanto, si se quieren más de dos iteraciones, el tamaño de paso se disminuye; y si se requieren menos de dos iteraciones el tamaño de paso se incrementa.

Aunque la estrategia anterior especifica el momento en que las modificaciones del tamaño de paso es el adecuado, no indica cómo habrá que cambiarlas. Ésta es una situación crítica, ya que los métodos de pasos múltiples requieren de varios puntos previos para calcular un nuevo punto. Una vez que se cambie el tamaño de paso, debe determinarse un nuevo conjunto de puntos. Un procedimiento es comenzar de nuevo los cálculos y usar el método de un paso para generar un nuevo conjunto de puntos de inicio.

Una estrategia más eficiente que utiliza la información existente consiste en aumentar y disminuir el tamaño de paso, mediante su duplicación o reducción a la mitad. Como se ilustra en la figura 26.6*b*, si se ha generado un número suficiente de valores previos, aumentar el tamaño de paso por duplicación será una tarea relativamente directa (figura 26.6*c*). Todo lo que se necesita es rastrear los subíndices de tal forma que los valores anteriores de *x* y de *y* sean los nuevos valores apropiados. La reducción del tamaño de paso a la mitad es más complicada, ya que algunos de los nuevos valores no estarán disponibles (figura 26.6*a*). Aunque se puede utilizar la interpolación polinomial del tipo que se desarrolló en el capítulo 18 para determinar estos valores intermedios.

En cualquier caso, la decisión de incorporar el control del tamaño de paso representa ventajas y desventajas entre la inversión inicial por la complejidad del programa contra los dividendos a largo plazo que se tendrán con el aumento en la eficiencia. En efecto, la magnitud e importancia del problema mismo tendrá un gran peso en dicha elección. Por fortuna, varios paquetes de software y bibliotecas tienen las rutinas de pasos múltiples que usted puede utilizar para obtener soluciones sin necesidad de programarlas desde las pruebas de escritorio. Mencionaremos algunas de éstas cuando revisemos los paquetes y las bibliotecas al final del capítulo 27.

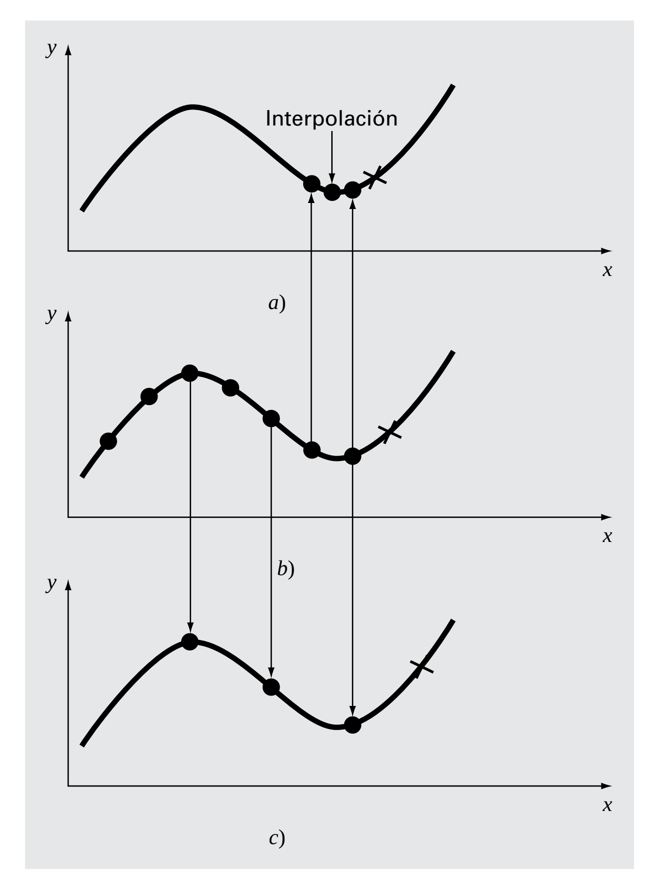


Figura 26.6
Una gráfica que muestra cómo la estrategia de disminuir a la mitad y duplicar el paso permite el uso de *b*) valores calculados previamente con un método de pasos múltiples de tercer orden. *a*) Disminuyendo a la mitad; *c*) duplicando.

26.2.3 Fórmulas de integración

El método de Heun sin autoinicio es característico de la mayoría de los métodos de pasos múltiples. Emplea una fórmula de integración abierta (el método del punto medio) para realizar una estimación inicial. Este paso predictor requiere un punto previo. Después, se aplica de manera iterativa una fórmula de integración cerrada (la regla del trapecio) para mejorar la solución.

Será evidente que una estrategia para mejorar los métodos de pasos múltiples será el uso de fórmulas de integración de orden superior como predictores y correctores. Por ejemplo, las fórmulas de Newton-Cotes de orden superior desarrolladas en el capítulo 21 se podrían utilizar para este propósito.

Antes de realizar una descripción de estos métodos de orden superior, revisaremos las fórmulas de integración más comunes en las que se basan. Como se mencionó antes, las primeras de éstas son las fórmulas de Newton-Cotes. No obstante, hay una segunda clase llamada fórmulas de Adams que también revisaremos y que a menudo se prefieren. Como se indica en la figura 26.7, la diferencia fundamental entre las fórmulas de Newton-Cotes y las de Adams tiene que ver con la manera en la cual se aplica la integral para obtener la solución. Como se muestra en la figura 26.7*a*, las fórmulas de Newton-Cotes estiman la integral en un intervalo generando varios puntos. Esta integral se usa entonces para proyectar la curva desde el inicio del intervalo hasta el final. En cambio, las fórmulas de Adams (figura 26.7*b*) emplean un conjunto de puntos de un intervalo para estimar únicamente la

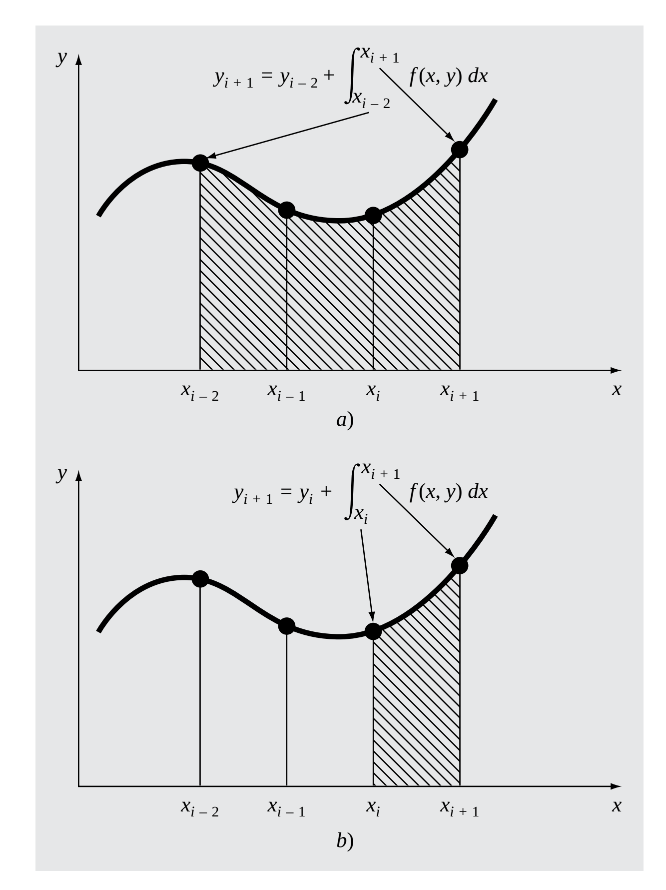


Figura 26.7
Ilustración de la diferencia fundamental entre las fórmulas de integración de Newton-Cotes y de Adams. *a*) Las fórmulas de Newton-Cotes utilizan una serie de puntos para obtener una estimación de la integral sobre varios segmentos. La estimación se usa después para proyectar la curva a través de todo el intervalo. *b*) Las fórmulas de Adams usan una serie de puntos para obtener la estimación de la integral de un solo segmento. La estimación después se utiliza para proyectar la curva a través del segmento.

integral del último segmento del intervalo. Esta integral se usa después para proyectar a través de este último segmento.

Fórmulas de Newton-Cotes Algunas de las fórmulas más conocidas para resolver ecuaciones diferenciales ordinarias se basan en ajustar un polinomio de interpolación de n-ésimo grado a $n + 1$ valores conocidos de y y, después esta ecuación, se utiliza para calcular la integral. Como se analizó antes en el capítulo 21, las fórmulas de integración de Newton-Cotes se basan en tal procedimiento y son de dos formas: abiertas y cerradas.

Fórmulas abiertas Para n valores igualmente espaciados, las fórmulas abiertas se pueden expresar en la forma de una solución para una EDO, como se hizo antes para la ecuación (26.19). La ecuación general para este propósito es

$$y_{i+1} = y_{i-n} + \int_{x_{i-n}}^{x_{i+1}} f_n(x)dx \tag{26.31}$$

donde $f_n(x)$ es un polinomio de interpolación de n-ésimo grado. La evaluación de la integral emplea la fórmula de integración abierta de Newton-Cotes de n-ésimo orden (tabla 21.4). Por ejemplo, si $n = 1$,

$$y_{i+1} = y_{i-1} + 2hf_i \tag{26.32}$$

donde f_i es una abreviatura de $f(x_i, y_i)$; es decir, la ecuación diferencial evaluada en x_i y y_i. Se hace referencia a la ecuación (26.32) como el método de punto medio y se utilizó antes como el predictor en el método de Heun sin autoinicio. Para $n = 2$,

$$y_{i+1} = y_{i-2} + \frac{3h}{2}(f_i + f_{i-1})$$

y para $n = 3$,

$$y_{i+1} = y_{i-3} + \frac{4h}{3}(2f_i - f_{i-1} + 2f_{i-2}) \tag{26.33}$$

La ecuación (26.33) se representa gráficamente en la figura 26.8*a*.

Fórmulas cerradas La forma cerrada se expresa de manera general como

$$y_{i+1} = y_{i-n+1} + \int_{x_{i-n+1}}^{x_{i+1}} f_n(x)\,dx \tag{26.34}$$

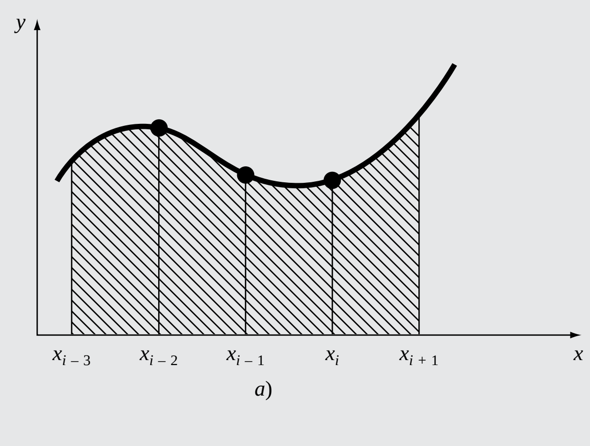
donde la integral se determina por una fórmula de integración cerrada de Newton-Cotes de n-ésimo orden (tabla 21.2). Por ejemplo, para $n = 1$,

$$y_{i+1} = y_i + \frac{h}{2}(f_i + f_{i+1})$$

que es equivalente a la regla del trapecio. Para $n = 2$,

$$y_{i+1} = y_{i-1} + \frac{h}{3}(f_{i-1} + 4f_i + f_{i+1}) \tag{26.35}$$

que es equivalente a la regla de Simpson 1/3. La ecuación (26.35) se representa en la figura 26.8*b*.



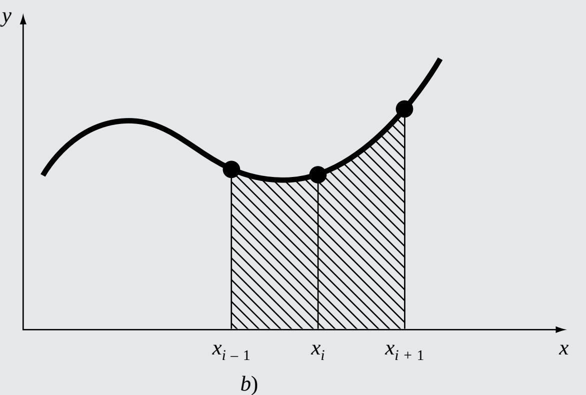


Figura 26.8
Representación gráfica de las fórmulas de integración de Newton-Cotes abierta y cerrada. *a*) La fórmula abierta de tercer orden [ecuación (26.33)] y *b*) la regla de Simpson 1/3 [ecuación (26.35)].

Fórmulas de Adams Los otros tipos de fórmulas de integración que se utilizan para resolver EDO son las fórmulas de Adams. Muchos algoritmos de uso generalizado para la solución de EDO por multipaso se basan en dichos métodos.

Fórmulas abiertas (Adams-Bashforth) Las fórmulas de Adams se deducen de varias formas. Una técnica consiste en escribir una expansión hacia adelante de la serie de Taylor alrededor de x_i:

$$y_{i+1} = y_i + f_i h + \frac{f_i'}{2}h^2 + \frac{f_i''}{6}h^3 + \cdots$$

que también se escribe como:

$$y_{i+1} = y_i + h\left(f_i + \frac{h}{2}f_i' + \frac{h^2}{3!}f_i'' + \cdots\right) \tag{26.36}$$

De la sección 4.1.3 recuerde que se puede usar una diferencia hacia atrás para aproximar la derivada:

$$f_i' = \frac{f_i - f_{i-1}}{h} + \frac{f_i''}{2}h + O(h^2)$$

que al sustituirse en la ecuación (26.36) da como resultado

$$y_{i+1} = y_i + h\left\{f_i + \frac{h}{2}\left[\frac{f_i - f_{i-1}}{h} + \frac{f_i''}{2}h + O(h^2)\right] + \frac{h^2}{6}f_i'' + \cdots\right\}$$

o, agrupando términos,

$$y_{i+1} = y_i + h\left(\frac{3}{2}f_i - \frac{1}{2}f_{i-1}\right) + \frac{5}{12}h^3 f_i'' + O(h^4) \tag{26.37}$$

Esta fórmula se conoce como la *fórmula abierta de Adams de segundo orden*. Las fórmulas abiertas de Adams también se denominan *fórmulas de Adams-Bashforth*. En consecuencia, la ecuación (26.37) se llama la *segunda fórmula de Adams-Bashforth*.

Es posible desarrollar fórmulas de Adams-Bashforth de orden superior sustituyendo aproximaciones por diferencias superiores en la ecuación (26.36). La fórmula abierta de Adams de *n*-ésimo orden en forma general se representa como:

$$y_{i+1} = y_i + h\sum_{k=0}^{n-1} \beta_k f_{i-k} + O(h^{n+1}) \tag{26.38}$$

Los coeficientes β_k se muestran en la tabla 26.1. La versión de cuarto orden se representa en la figura 26.9*a*. Observe que la versión de primer orden es el método de Euler.

Fórmulas cerradas (de Adams-Moulton) Una serie de Taylor hacia atrás alrededor de x_{i+1} se escribe como:

$$y_i = y_{i+1} - f_{i+1}h + \frac{f_{i+1}'}{2}h^2 - \frac{f_{i+1}''}{3!}h^3 + \cdots$$

Tabla 26.1 Coeficientes y error de truncamiento para los predictores de Adams-Bashforth.

Orden	β_0	β_1	β_2	β_3	β_4	β_5	Error de truncamiento local
1	1						$\frac{1}{2}h^2 f'(\xi)$
2	3/2	–1/2					$\frac{5}{12}h^3 f''(\xi)$
3	23/12	–16/12	5/12				$\frac{9}{24}h^4 f^{(3)}(\xi)$
4	55/24	–59/24	37/24	–9/24			$\frac{251}{720}h^5 f^{(4)}(\xi)$
5	1 901/720	–2 774/720	2 616/720	–1 274/720	251/720		$\frac{475}{1440}h^6 f^{(5)}(\xi)$
6	4 277/720	–7 923/720	9 982/720	–7 298/720	2 877/720	–475/720	$\frac{19\,087}{60\,480}h^7 f^{(6)}(\xi)$

Al resolver para y_{i+1} se obtiene:

$$y_{i+1} = y_i + h\left(f_{i+1} - \frac{h}{2} f'_{i+1} + \frac{h^2}{6} f''_{i+1} + \cdots \right. \tag{26.39}$$

Se puede usar una diferencia para aproximar la primera derivada:

$$f'_{i+1} = \frac{f_{i+1} - f_i}{h} + \frac{f''_{i+1}}{2} h + O(h^2)$$

que al sustituirse en la ecuación (26.39), y agrupando términos, da

$$y_{i+1} = y_i + h\left(\frac{1}{2} f_{i+1} + \frac{1}{2} f_i \right) - \frac{1}{12} h^3 f''_{i+1} - O(h^4)$$

Esta fórmula se conoce como la *fórmula cerrada de Adams de segundo orden* o la *segunda fórmula de Adams-Moulton*. Observe también que es la regla del trapecio.

La fórmula cerrada de Adams de *n*-ésimo orden generalmente se escribe como:

$$y_{i+1} = y_i + h \sum_{k=0}^{n-1} \beta_k f_{i+1-k} + O(h^{n+1})$$

Los coeficientes β_k se muestran en la tabla 26.2. El método de cuarto orden se ilustra en la figura 26.9*b*.

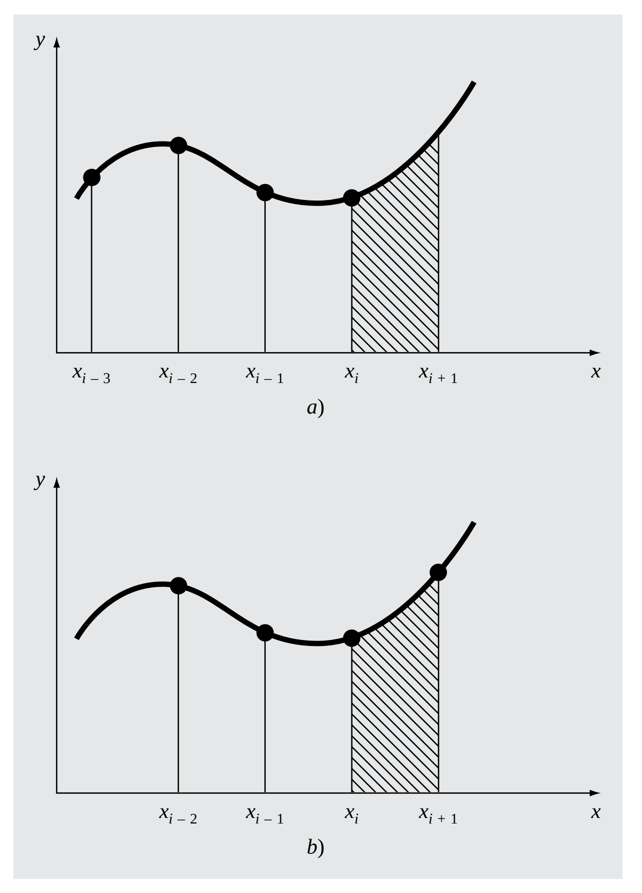


Figura 26.9
Representación gráfica de las fórmulas de integración de Adams abierta y cerrada. *a*) La cuarta fórmula de Adams-Bashforth abierta y *b*) la cuarta fórmula de Adams-Moulton cerrada.

26.2.4 Métodos de pasos múltiples de orden superior

Ahora que formalmente desarrollamos las fórmulas de integración de Newton-Cotes y de Adams, podemos utilizarlas para deducir métodos de pasos múltiples de orden superior. Como ocurrió con el método de Heun sin autoinicio, las fórmulas de integración se aplican conjuntamente como métodos predictor-corrector. Además, si las fórmulas abiertas y cerradas tienen errores de truncamiento local del mismo orden, es posible incorporar modificadores del tipo que se presenta en la figura 26.5 para mejorar la exactitud y permitir el control del tamaño de paso. El cuadro 26.1 ofrece ecuaciones generales para estos modificadores. En la siguiente sección presentamos dos de los procedimientos multipaso de orden superior más comunes: el método de Milne y el método de Adams de cuarto orden.

Tabla 26.2 Coeficientes y error de truncamiento de los correctores de Adams-Moulton.

Orden	β_0	β_1	β_2	β_3	β_4	β_5	Error de truncamiento local
2	1/2	1/2					$-\frac{1}{12}h^3f''(\xi)$
3	5/12	8/12	–1/12				$-\frac{1}{24}h^4f^{(3)}(\xi)$
4	9/24	19/24	–5/24	1/24			$-\frac{19}{720}h^5f^{(4)}(\xi)$
5	251/720	646/720	–264/720	106/720	–19/720		$-\frac{27}{1\,440}h^6f^{(5)}(\xi)$
6	475/1 440	1 427/1 440	–798/1 440	482/1 440	–173/1 440	27/1 440	$-\frac{863}{60\,480}h^7f^{(6)}(\xi)$

Cuadro 26.1 Deducción de relaciones generales para modificadores

La relación entre el valor verdadero, la aproximación y el error de un predictor se representa en forma general como

$$\text{Valor verdadero} = y_{i+1}^0 + \frac{\eta_p}{\delta_p}h^{n+1}y^{(n+1)}(\xi_p) \qquad \text{(C26.1.1)}$$

donde η_p y δ_p = numerador y denominador, respectivamente, de la constante del error de truncamiento de un predictor, ya sea de Newton-Cotes abierto (tabla 21.4) o de Adams-Bashforth (tabla 26.1) y n es el orden.

Se desarrolla una relación similar para el corrector:

$$\text{Valor verdadero} = y_{i+1}^m - \frac{\eta_c}{\delta_c}h^{n+1}y^{(n+1)}(\xi_c) \qquad \text{(C26.1.2)}$$

donde η_c y δ_c = numerador y denominador, respectivamente, de la constante del error de truncamiento de un corrector, ya sea de Newton-Cotes cerrado (tabla 21.2) o de Adams-Moulton (tabla 26.2). Como en la deducción de la ecuación (26.24), la ecuación (C26.1.1) se resta de la ecuación (C26.1.2) para obtener

$$0 = y_{i+1}^m - y_{i+1}^0 - \frac{\eta_c + \eta_p\delta_c/\delta_p}{\delta_c}h^{n+1}y^{(n+1)}(\xi) \qquad \text{(C26.1.3)}$$

Ahora, dividiendo la ecuación entre $\eta_c + \eta_p\,\delta_c/\delta_p$, multiplicando el último término por δ_p/δ_p y reordenando, se obtiene una estimación del error de truncamiento local del corrector:

$$E_c \cong -\frac{\eta_c\delta_p}{\eta_c\delta_p + \eta_p\delta_c}(y_{i+1}^m - y_{i+1}^0) \qquad \text{(C26.1.4)}$$

Para el modificador del predictor, la ecuación (C26.1.3) se despeja en el paso anterior:

$$h^n y^{(n+1)}(\xi) = -\frac{\delta_c\delta_p}{\eta_c\delta_p + \eta_p\delta_c}(y_i^0 - y_i^m)$$

que podrá sustituirse en el término del error de la ecuación (C26.1.1) para tener

$$E_p = \frac{\eta_p\delta_c}{\eta_c\delta_p + \eta_p\delta_c}(y_i^m - y_i^0) \qquad \text{(C26.1.5)}$$

Las ecuaciones (C26.1.4) y (C26.1.5) son versiones generales de modificadores que se utilizan para mejorar algoritmos de pasos múltiples. Por ejemplo, el método de Milne tiene η_p = 14, δ_p = 45, η_c = 1, δ_c = 90. Sustituyendo estos valores en las ecuaciones (C26.1.4) y (C26.1.5) se obtienen las ecuaciones (26.43) y (26.42), respectivamente. Podrán desarrollarse modificadores similares para otros pares de fórmulas abiertas y cerradas que tengan errores de truncamiento local del mismo orden.

Método de Milne El método de Milne es el método de pasos múltiples más común, basado en las fórmulas de integración de Newton-Cotes. Éste utiliza la fórmula abierta de Newton-Cotes de tres puntos como un predictor:

$$y_{i+1}^0 = y_{i-3}^m + \frac{4h}{3}(2f_i^m - f_{i-1}^m + 2f_{i-2}^m) \tag{26.40}$$

y la fórmula cerrada de Newton-Cotes de tres puntos (regla de Simpson 1/3) como corrector:

$$y_{i+1}^j = y_{i-1}^m + \frac{h}{3}(f_{i-1}^m + 4f_i^m + f_{i+1}^{j-1}) \tag{26.41}$$

donde j es un índice que representa el número de iteraciones del modificador. El predictor y los modificadores del corrector para el método de Milne se desarrollan a partir de las fórmulas del cuadro 26.1 y de los coeficientes del error en las tablas 21.2 y 21.4:

$$E_p = \frac{28}{29}(y_i^m - y_i^0) \tag{26.42}$$

$$E_c \cong -\frac{1}{29}(y_{i+1}^m - y_{i+1}^0) \tag{26.43}$$

EJEMPLO 26.5 Método de Milne

Planteamiento del problema Con el método de Milne integre $y' = 4e^{0.8x} - 0.5y$ desde $x = 0$ hasta $x = 4$ usando un tamaño de paso de 1. La condición inicial en $x = 0$, es $y = 2$. Como tratamos con un método de pasos múltiples se necesita de puntos previos. En una aplicación real, se usaría un método de un paso tal como un RK de cuarto orden para calcular los puntos requeridos. En este ejemplo, usaremos la solución analítica [recuerde la ecuación (E25.5.1) del ejemplo 25.5] obteniéndose los valores exactos para $x_{i-3} = -3$, $x_{i-2} = -2$ y $x_{i-1} = -1$ de $y_{i-3} = -4.547302$, $y_{i-2} = -2.306160$ y $y_{i-1} = -0.3929953$, respectivamente.

Solución El predictor [ecuación (26.40)] se usa para calcular el valor en $x = 1$:

$$y_1^0 = -4.54730 + \frac{4(1)}{3}[2(3) - 1.99381 + 2(1.96067)] = 6.02272 \qquad \varepsilon_t = 2.8\%$$

El corrector [ecuación (26.41)] se emplea después para calcular

$$y_1^1 = -0.3929953 + \frac{1}{3}[1.99381 + 4(3) + 5.890802] = 6.235210 \qquad \varepsilon_t = -0.66\%$$

Este resultado puede sustituirse en la ecuación (26.41) para corregir la estimación en forma iterativa. El proceso converge a un valor final corregido de 6.204855 ($\varepsilon_t = -0.17\%$).

Este valor es más exacto que la estimación de 6.360865 ($\varepsilon_t = -2.68\%$) obtenida antes con el método de Heun sin autoinicio (los ejemplos 26.2 a 26.4). Los resultados en los siguientes pasos son $y(2) = 14.86031$ ($\varepsilon_t = -0.11\%$), $y(3) = 33.72426$ ($\varepsilon_t = -0.14\%$) y $y(4) = 75.43295$ ($\varepsilon_t = -0.12\%$).

Como en el ejemplo anterior, el método de Milne generalmente da resultados de gran exactitud. Sin embargo, hay ciertos casos donde su desempeño es pobre (véase Ralston y Rabinowitz, 1978). Antes de examinar tales casos, describiremos otro procedimiento multipaso de orden superior: el método de Adams de cuarto orden.

Método de Adams de cuarto orden Un método común de pasos múltiples basado en las fórmulas de integración de Adams utiliza la fórmula de Adams-Bashforth de cuarto orden (tabla 26.1) como predictor:

$$y_{i+1}^0 = y_i^m + h\left(\frac{55}{24}f_i^m - \frac{59}{24}f_{i-1}^m + \frac{37}{24}f_{i-2}^m - \frac{9}{24}f_{i-3}^m\right) \tag{26.44}$$

y la fórmula de Adams-Moulton de cuarto orden (tabla 26.2) como corrector:

$$y_{i+1}^{j} = y_i^m + h\left(\frac{9}{24}f_{i+1}^{j-1} + \frac{19}{24}f_i^m - \frac{5}{24}f_{i-1}^m + \frac{1}{24}f_{i-2}^m\right) \tag{26.45}$$

Los modificadores del predictor y del corrector para el método de Adams de cuarto orden se desarrollan a partir de las fórmulas del cuadro 26.1 y de los coeficientes de error en las tablas 26.1 y 26.2 como sigue:

$$E_p = \frac{251}{270}(y_i^m - y_i^0) \tag{26.46}$$

$$E_c = -\frac{19}{270}(y_{i+1}^m - y_{i+1}^0) \tag{26.47}$$

EJEMPLO 26.6 Método de Adams de cuarto orden

Planteamiento del problema Con el método de Adams de cuarto orden resuelva el mismo problema que en el ejemplo 26.5.

Solución El predictor [ecuación (26.44)] se utiliza para calcular el valor en $x = 1$.

$$y_1^0 = 2 + 1\left(\frac{55}{24}3 - \frac{59}{24}1.993814 + \frac{37}{24}1.960667 - \frac{9}{24}2.6365228\right) = 6.007539$$

$$\varepsilon_t = 3.1\%$$

el cual es comparable al resultado que se obtiene usando el método de Milne, aunque menos exacto. El corrector [ecuación (26.45)] se emplea después para calcular

$$y_1^1 = 2 + 1\left(\frac{9}{24}5.898394 + \frac{19}{24}3 - \frac{5}{24}1.993814 + \frac{1}{24}1.960666\right) = 6.253214$$

$$\varepsilon_t = -0.96\%$$

que también es comparable, aunque menos exacto que el resultado con el método de Milne. Este resultado se sustituye en la ecuación (26.45) para corregir de manera iterativa el estimado. El proceso converge a un valor corregido final de 6.214424 ($\varepsilon_t = 0.32\%$), que es un resultado exacto, pero también inferior al obtenido con el método de Milne.

Estabilidad de los métodos de pasos múltiples La mejor exactitud del método de Milne mostrada en los ejemplos 26.5 y 26.6 podría anticiparse considerando los términos del error en los predictores [ecuaciones (26.42) y (26.46)] y en los correctores [ecuaciones (26.43) y (26.47)]. Los coeficientes para el método de Milne, 14/45 y 1/90, son más pequeños que los de Adams de cuarto orden, 251/720 y 19/720. Además, el método de Milne emplea menos evaluaciones de la función para alcanzar mejores estimaciones. Los anteriores resultados podrían llevarnos a la conclusión de que el método de Milne es superior y, por lo tanto, preferible a los de Adams de cuarto orden. Aunque esta conclusión se cumple en muchos casos, hay ocasiones en las que el método de Milne se desempeña en forma inaceptable. Tal comportamiento se muestra en el siguiente ejemplo.

EJEMPLO 26.7 Estabilidad de los métodos de Milne y de Adams de cuarto orden

Planteamiento del problema Use los métodos de Milne y de Adams de cuarto orden para resolver

$$\frac{dy}{dx} = -y$$

con la condición inicial $x = 0$, $y = 1$. Resuelva esta ecuación desde $x = 0$ hasta $x = 10$ usando un tamaño de paso $h = 0.5$. Observe que la solución analítica es $y = e^{-x}$.

Solución Los resultados, como se resumen en la figura 26.10, indican problemas con el método de Milne. Poco después del inicio de los cálculos, los errores empiezan a crecer y a oscilar en signo. En $x = 10$ el error relativo ha crecido a 2 831% y el valor predicho mismo comienza a oscilar en signo.

En cambio, los resultados con el método de Adams serán mucho más aceptables. Aunque el error también crezca, lo hará lentamente. Además, las discrepancias no mostrarán los violentos cambios de signo que muestra el método de Milne.

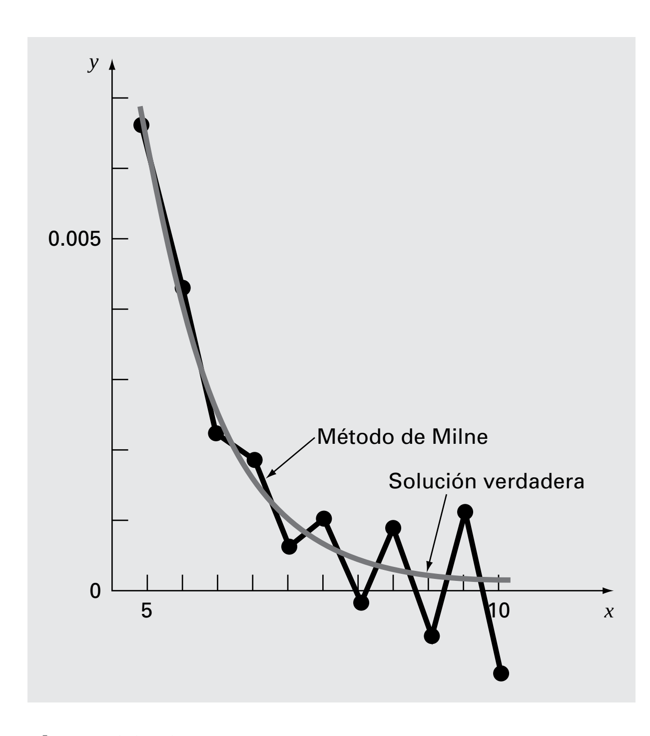


Figura 26.10
Representación gráfica de la inestabilidad del método de Milne.

Al inaceptable comportamiento manifestado por el método de Milne en el ejemplo anterior se le conoce como inestabilidad. Aunque no siempre ocurre, tal posibilidad nos lleva a la conclusión de que deberá evitarse el procedimiento de Milne. Así, en general se prefiere el método de Adams de cuarto orden.

La inestabilidad del método de Milne se debe al corrector. En consecuencia, se han realizado intentos para rectificar el defecto al desarrollar correctores estables. Una alternativa usada comúnmente que emplea este procedimiento es el *método de Hamming*, el cual utiliza el predictor de Milne y un corrector estable:

$$y_{i+1}^{j} = \frac{9y_i^m - y_{i-2}^m + 3h(y_{i+1}^{j-1} + 2f_i^m - f_{i-1}^m)}{8}$$

que tiene un error de truncamiento local:

$$E_c = \frac{1}{40}h^5 y^{(4)}(\xi_c)$$

El método de Hamming también implica modificadores de la forma:

$$E_p = \frac{9}{121}(y_i^m - y_i^0)$$

$$E_c = -\frac{112}{121}(y_{i+1}^m - y_{i+1}^0)$$

El lector encontrará información adicional sobre éste y otros métodos de pasos múltiples en muchas fuentes (Hamming, 1973; Lapidus y Seinfield, 1971).

PROBLEMAS

26.1 Dada

$$\frac{dy}{dx} = -200\,000y + 200\,000e^{-x} - e^{-x}$$

a) Estime el tamaño de paso requerido para mantener la estabilidad con el uso del método de Euler explícito.
b) Si $y(0) = 0$, utilice el método de Euler implícito para obtener la solución desde $t = 0$ hasta 2, con un tamaño de paso de 0.1.

26.2 Dado que

$$\frac{dy}{dt} = 30(\cos t - y) + 3\,\text{sen}\,t$$

Si $y(0) = 1$, emplee el método de Euler implícito para obtener una solución de $t = 0$ a 4, con un tamaño de paso de 0.4.

26.3 Dadas

$$\frac{dx_1}{dt} = 1\,999x_1 + 2\,999x_2$$

$$\frac{dx_2}{dt} = -2\,000x_1 - 3\,000x_2$$

Si $x_1(0) = x_2(0) = 1$, obtenga una solución de $t = 0$ a 0.2, con un tamaño de paso de 0.05, con los métodos de Euler *a*) explícito y *b*) implícito.

26.4 Resuelva el siguiente problema de valor inicial sobre el intervalo de $t = 2$ a 3:

$$\frac{dy}{dt} = -0.4y + e^{-2t}$$

Use el método de Heun sin autoinicio, con un tamaño de paso de 0.5 y condiciones iniciales de $y(1.5) = 5.800007$ y $y(2.0) = 4.762673$. Itere el corrector hasta $\varepsilon_s = 0.1\%$. Calcule los errores relativos verdaderos porcentuales ε_t para sus resultados, con base en la solución analítica.

26.5 Repita el problema 26.4, pero use el método de Adams de cuarto orden. [Observe que $y(0.5) = 8.46909$ y $y(1.0) = 7.037566$]. Itere el corrector hasta $\varepsilon_s = 0.01\%$.

26.6 Resuelva el problema siguiente de valor inicial, de $t = 4$ a 5:

$$\frac{dy}{dt} = -\frac{2y}{t}$$

Use un tamaño de paso de 0.5 y valores iniciales de $y(2.5) = 0.48$, $y(3) = 0.333333$, $y(3.5) = 0.244898$, y $y(4) = 0.1875$. Obtenga sus soluciones con las técnicas siguientes: *a*) método de Heun sin autoinicio ($\varepsilon_s = 1\%$) y *b*) método de Adams de cuarto orden ($\varepsilon_s = 0.01\%$). [Nota: las respuestas correctas que se obtienen de forma analítica son $y(4.5) = 0.148148$ y $y(5) = 0.12$.] Calcule los errores relativos porcentuales verdaderos ε_t de sus resultados.

26.7 Resuelva el problema que sigue de valor inicial de $x = 0$ a $x = 0.75$:

$$\frac{dy}{dx} = yx^2 - y$$

Utilice el método de Heun sin autoinicio con tamaño de paso de 0.25. Si $y(0) = 1$, emplee el método de RK de cuarto orden con tamaño de paso de 0.25 para predecir el valor de inicio en $y(0.25)$.

26.8 Solucione el problema siguiente de valor inicial, de $t = 1.5$ a $t = 2.5$

$$\frac{dy}{dt} = \frac{-2y}{1+t}$$

Use el método de Adams de cuarto orden. Emplee un tamaño de paso de 0.5 y el método de RK de cuarto orden para pronosticar los valores de inicio si $y(0) = 2$.

26.9 Desarrolle un programa para el método de Euler implícito para una EDO lineal. Pruébelo con la repetición del problema 26.1*b*.

26.10 Desarrolle un programa para el método de Euler implícito para un par de EDO lineales. Pruébelo con la solución de la ecuación (26.6).

26.11 Desarrolle un programa amigable para el usuario para el método de Heun sin autoinicio con modificador predictor. Emplee el método de RK de cuarto orden para calcular valores de inicio. Pruebe el programa con la repetición del ejemplo 26.4.

26.12 Use el programa desarrollado en el problema 26.11 para resolver el problema 26.7.

26.13 Considere la barra delgada de longitud l que se mueve en el plano x-y, como se ilustra en la figura P26.13. La barra se fija en uno de sus extremos con un perno y con una masa en el otro. Observe que $g = 9.81$ m/s² y $l = 0.5$ m. Este sistema se puede resolver usando:

$$\ddot{\theta} - \frac{g}{l}\theta = 0$$

Sea $\theta(0) = 0$ y $\dot{\theta}(0) = 0.25$ rad/s. Resuelva con cualquiera de los métodos que se estudiaron en este capítulo. Grafique el ángulo *versus* el tiempo, y la velocidad angular *versus* el tiempo. (Recomendación: descomponga la EDO de segundo orden.)

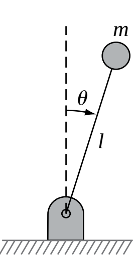


Figura P26.13

26.14 Dada la EDO de primer orden:

$$\frac{dx}{dt} = -700x - 1\,000e^{-t}$$
$$x(t = 0) = 4$$

Resuelva esta ecuación diferencial rígida con algún método numérico, en el periodo $0 \le t \le 5$. También resuélvala en forma analítica y grafique las soluciones analítica y numérica tanto para la fase de transición rápida como lenta de la escala temporal.

26.15 Se considera que la siguiente EDO de segundo orden es rígida:

$$\frac{d^2y}{dx^2} = -1\,001\frac{dy}{dx} - 1\,000y$$

Resuelva esta ecuación diferencial en forma *a*) analítica y *b*) numérica, de $x = 0$ a 5. Para el inciso *b*) utilice un enfoque implícito con $h = 0.5$. Observe que las condiciones iniciales son $y(0) = 1$ y $y'(0) = 0$. Muestre los dos resultados gráficamente.

26.16 Resuelva la ecuación diferencial siguiente, de $t = 0$ a 1

$$\frac{dy}{dt} = -10y$$

con la condición inicial $y(0) = 1$. Use las técnicas siguientes para obtener sus soluciones: *a*) analítica, *b*) método de Euler explícito y *c*) método de Euler implícito. Para los incisos *b*) y *c*) use $h = 0.1$ y 0.2. Grafique sus resultados.

CAPÍTULO 27

Problemas de valores en la frontera y de valores propios

De nuestro análisis al inicio de la parte siete, recuerde que una ecuación diferencial ordinaria se acompaña de condiciones auxiliares. Estas condiciones se utilizan para evaluar las constantes de integración que resultan durante la solución de la ecuación. Para una ecuación de n-ésimo orden, se requieren n condiciones. Si todas las condiciones se especifican para el mismo valor de la variable independiente, entonces se trata de un *problema de valor inicial* (figura 27.1*a*). Hasta aquí, el material de la parte siete se ha dedicado a este tipo de problema.

Hay otra aplicación en la cual las condiciones no se conocen para un solo punto, sino, más bien, se conocen en diferentes valores de la variable independiente. Debido a que estos valores se especifican en los puntos extremos o frontera de un sistema, se les conoce como *problemas de valores en la frontera* (figura 27.1*b*). Muchas aplicaciones importantes en ingeniería son de esta clase. En el presente capítulo analizamos dos procedimientos generales para obtener su solución: el método de disparo y la aproximación en diferencias finitas. Además, presentamos técnicas para abordar un tipo especial de problema de valores en la frontera: la determinación de valores propios (valores característicos o eigenvalores). Por supuesto, los valores propios también tienen muchas aplicaciones que van más allá de las relacionadas con los problemas de valores en la frontera.

27.1 MÉTODOS GENERALES PARA PROBLEMAS DE VALORES EN LA FRONTERA

Se puede utilizar la conservación de la energía para desarrollar un balance de calor para una barra larga y delgada (figura 27.2). Si la barra no está aislada en toda su longitud y el sistema se encuentra en estado estacionario, la ecuación resultante es

$$\frac{d^2T}{dx^2} + h'(T_a - T) = 0 \tag{27.1}$$

donde h' es un coeficiente de transferencia de calor (m^{-2}) que parametriza la velocidad con que se disipa el calor en el medio ambiente, y T_a es la temperatura del medio ambiente (°C).

Para obtener una solución de la ecuación (27.1) se deben tener condiciones de frontera adecuadas. Un caso simple es aquel donde los valores de las temperaturas en los extremos de la barra se mantienen fijos. Estos valores se expresan en forma matemática como

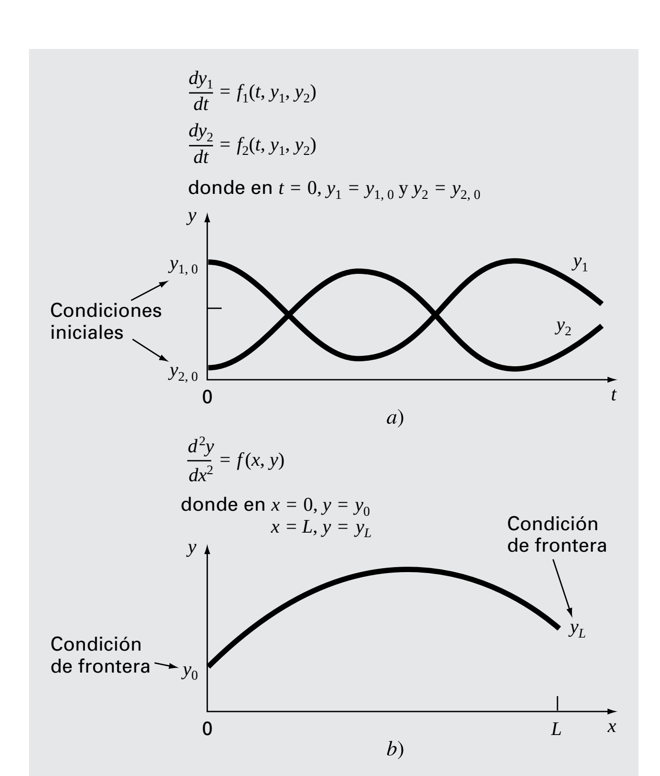


Figura 27.1
Problemas de valor inicial contra problemas de valores en la frontera. *a*) Un problema de valor inicial donde todas las condiciones se especifican para el mismo valor de la variable independiente. *b*) Un problema con valores en la frontera donde las condiciones se especifican para diferentes valores de la variable independiente.

$$T(0) = T_1$$

$$T(L) = T_2$$

Con estas condiciones, la ecuación (27.1) se puede resolver de manera analítica usando el cálculo. Para una barra de 10 metros con $T_a = 20$, $T_1 = 40$, $T_2 = 200$ y $h' = 0.01$, la solución es

$$T = 73.4523e^{0.1x} - 53.4523e^{-0.1x} + 20 \qquad (27.2)$$

En las siguientes secciones se resolverá el mismo problema usando procedimientos numéricos.

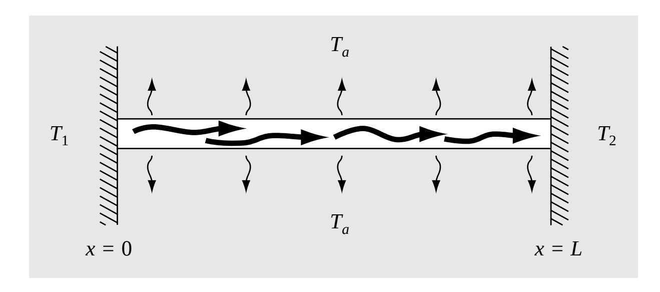


Figura 27.2
Una barra uniforme no aislada colocada entre dos cuerpos de temperatura constante, pero diferente. En este caso, $T_1 > T_2$ y $T_2 > T_a$.

27.1.1 El método de disparo

El *método de disparo* se basa en convertir el problema de valor en la frontera en un problema de valor inicial equivalente. Posteriormente se aplica un procedimiento de prueba y error para resolver la versión del valor inicial. El método se ilustrará con un ejemplo.

EJEMPLO 27.1 El método de disparo

Planteamiento del problema Utilice el método de disparo para resolver la ecuación (27.1), para una barra de 10 metros, $h' = 0.01\ \text{m}^{-2}$, $T_a = 20$ y las condiciones de frontera

$$T(0) = 40 \qquad T(10) = 200$$

Solución Usando el mismo procedimiento que se empleó para transformar la ecuación (PT7.2) en las ecuaciones (PT7.3) a (PT7.6), la ecuación diferencial de segundo orden se expresa como dos EDO de primer orden:

$$\frac{dT}{dx} = z \qquad (E27.1.1)$$

$$\frac{dz}{dx} = h'(T - T_a) \qquad (E27.1.2)$$

Para resolver estas ecuaciones, se requiere un valor inicial para z. En el método de disparo, proponemos un valor inicial, digamos, $z(0) = 10$. La solución se obtiene integrando las ecuaciones (E27.1.1) y (E27.1.2) simultáneamente. Por ejemplo, utilizando un método RK de cuarto orden con un tamaño de paso de 2, obtenemos un valor en el extremo del intervalo, $T(10) = 168.3797$ (figura 27.3*a*), el cual difiere de la condición de frontera, $T(10) = 200$. Por lo tanto, debemos realizar otra suposición, $z(0) = 20$, y efectuar de nuevo el cálculo. Esta vez, se obtiene el resultado de $T(10) = 285.8980$ (figura 27.3*b*).

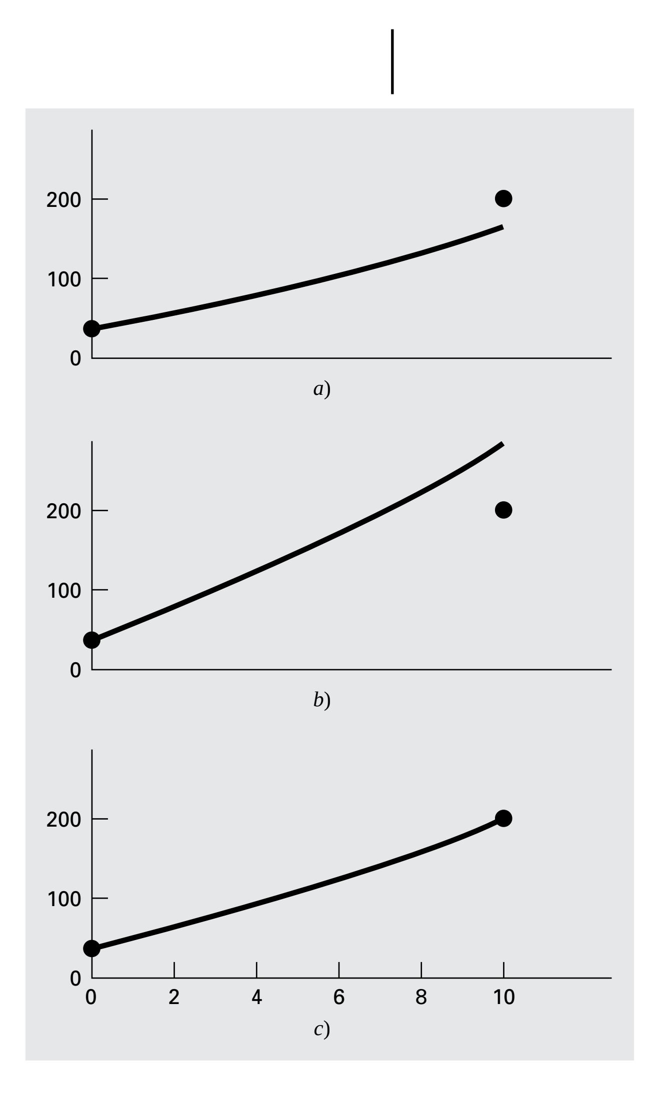


Figura 27.3
El método de disparo: *a*) el primer "disparo", *b*) el segundo "disparo" y *c*) el "tiro" final exacto.

Ahora, como la EDO original es lineal, los valores

$z(0) = 10 \qquad T(10) = 168.3797$

y

$z(0) = 20 \qquad T(10) = 285.8980$

están relacionados linealmente. Así, pueden utilizarse para calcular el valor de $z(0)$ que da $T(10) = 200$. Se emplea una fórmula de interpolación lineal [recuerde la ecuación (18.2)] para tal propósito:

$$z(0) = 10 + \frac{20 - 10}{285.8980 - 168.3797}(200 - 168.3797) = 12.6907$$

Este valor se utiliza después para determinar la solución correcta, como se ilustra en la figura 27.3*c*.

Problemas no lineales de dos puntos Para problemas de valores en la frontera no lineales, la interpolación lineal o extrapolación por medio de dos puntos de solución no necesariamente dará como resultado una estimación exacta de la condición de frontera requerida para obtener una solución exacta. Una alternativa consiste en realizar tres aplicaciones del método de disparo y usar un polinomio de interpolación cuadrática para estimar la condición de frontera adecuada. No obstante, es poco probable que este procedimiento ofrezca la respuesta exacta, y se necesitarán iteraciones adicionales para llegar a la solución.

Otro procedimiento para un problema no lineal implica reformularlo como un problema de raíces. Recuerde que la forma general de un problema de raíces consiste en encontrar el valor de x que haga la función $f(x) = 0$. Ahora, usemos el ejemplo 27.1 para entender cómo se plantea en esta forma el método de disparo.

Primero, reconocemos que la solución del par de ecuaciones diferenciales es también una “función”, en el sentido que suponemos una condición en el extremo izquierdo de la barra, z_0, y la integración nos da una predicción de la temperatura en el extremo derecho, T_{10}. Así, se considera la integración como:

$$T_{10} = f(z_0)$$

Es decir, la integración representa un proceso por medio del cual un supuesto de z_0 dará una predicción de T_{10}. Visto de esta manera, sabemos que lo que deseamos es el valor de z_0 que proporcione un valor específico de T_{10}. Si, como en el ejemplo, deseamos $T_{10} = 200$, planteamos el problema como sigue:

$$200 = f(z_0)$$

Llevando el 200, que es nuestro objetivo, al lado derecho de la ecuación, genera una nueva función, $g(z_0)$, que representa la diferencia entre lo que tenemos, $f(z_0)$, y lo que buscamos, 200.

$$g(z_0) = f(z_0) - 200$$

Si llevamos esta nueva función a cero, obtendremos la solución. El siguiente ejemplo ilustra el procedimiento.

EJEMPLO 27.2 El método de disparo para problemas no lineales

Planteamiento del problema Aunque sirvió para nuestros propósitos plantear un problema sencillo de valor en la frontera, nuestro modelo para la barra en la ecuación (27.1) no fue muy realista. Debido a que la barra perderá calor por mecanismos que son no lineales, como la radiación.

Suponga que la siguiente EDO no lineal se utiliza para modelar la temperatura de la barra caliente:

$$\frac{d^2T}{dx^2} + h''(T_a - T)^4 = 0$$

donde $h'' = 5 \times 10^{-8}$. Ahora, aunque todavía no es una muy buena representación de la transferencia del calor, esta ecuación es suficientemente clara para permitirnos ilustrar cómo se utiliza el método de disparo para resolver un problema de valor en la frontera no lineal de dos puntos. Las condiciones restantes del problema son las que se especifican en el ejemplo 27.1.

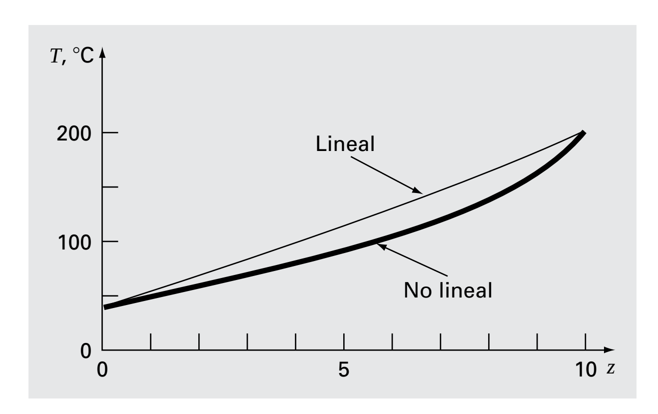


Figura 27.4
El resultado de usar el método de disparo para resolver un problema no lineal.

Solución La ecuación diferencial de segundo orden se expresa como dos EDO de primer orden:

$$\frac{dT}{dx} = z$$

$$\frac{dz}{dx} = h''(T - T_a)^4$$

Ahora, se integran estas ecuaciones usando cualquiera de los métodos que se describen en los capítulos 25 y 26. Utilizamos la versión con tamaño de paso constante del método de RK de cuarto orden del capítulo 25. Donde implementamos este procedimiento como una función macro de Excel escrita en Visual BASIC. La función integró las ecuaciones partiendo de un valor inicial para $z(0)$ y dio como resultado la temperatura en $x = 10$. La diferencia entre este valor y el objetivo de 200 se introdujo luego en una celda de la hoja de cálculo. El Solver de Excel se utilizó después para ajustar el valor de $z(0)$ hasta que la diferencia fuera cero.

El resultado se muestra en la figura 27.4 junto con el caso lineal original. Como se esperaba, el caso no lineal está más "curvado" que el modelo lineal. Lo anterior se debe al término a la cuarta potencia en la relación de la transferencia del calor.

El método de disparo se vuelve difícil para ecuaciones de orden superior, donde la necesidad de suponer dos o más condiciones vuelve el procedimiento más difícil. Por tales razones, se dispone de métodos alternativos que se describen a continuación.

27.1.2 Métodos de diferencias finitas

Las alternativas más comunes al método de disparo son los *métodos por diferencias finitas*, en las cuales, las diferencias divididas finitas sustituyen a las derivadas en la ecuación original. Así, una ecuación diferencial lineal se transforma en un conjunto de ecuaciones algebraicas simultáneas que pueden resolverse utilizando los métodos de la parte tres.

En el caso de la figura 27.2, la aproximación en diferencias divididas finitas para la segunda derivada es (recuerde la figura 23.3)

$$\frac{d^2T}{dx^2} = \frac{T_{i+1} - 2T_i + T_{i-1}}{\Delta x^2}$$

Esta aproximación se sustituye en la ecuación (27.1) para dar

$$\frac{T_{i+1} - 2T_i + T_{i-1}}{\Delta x^2} - h'(T_i - T_a) = 0$$

Agrupando términos se tiene

$$-T_{i-1} + (2 + h'\Delta x^2)T_i - T_{i+1} = h'\Delta x^2 T_a \tag{27.3}$$

Esta ecuación es válida para cada uno de los nodos interiores de la barra. Los nodos interiores primero y último, T_{i-1} y T_{i+1}, respectivamente, se especifican por las condiciones de frontera. Por lo tanto, el conjunto resultante de ecuaciones algebraicas lineales será tridiagonal. Como tal, se resuelve con los algoritmos eficientes de que se dispone para estos sistemas (sección 11.1).

EJEMPLO 27.3 Aproximación por diferencias finitas de problemas con valores en la frontera

Planteamiento del problema Use el procedimiento por diferencias finitas para resolver el mismo problema que en el ejemplo 27.1.

Solución Empleando los parámetros del ejemplo 27.1, se escribe la ecuación (27.3) para la barra mostrada en la figura 27.2. El empleo de cuatro nodos interiores con un segmento de longitud $\Delta x = 2$ metros da como resultado las siguientes ecuaciones:

$$\begin{bmatrix} 2.04 & -1 & 0 & 0 \\ -1 & 2.04 & -1 & 0 \\ 0 & -1 & 2.04 & -1 \\ 0 & 0 & -1 & 2.04 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{Bmatrix} = \begin{Bmatrix} 40.8 \\ 0.8 \\ 0.8 \\ 200.8 \end{Bmatrix}$$

de las cuales se obtienen

$$\{T\}^T = \lfloor 65.9698 \quad 93.7785 \quad 124.5382 \quad 159.4795 \rfloor$$

La tabla 27.1 ofrece una comparación entre la solución analítica [ecuación (27.2)] y las soluciones numéricas obtenidas en los ejemplos 27.1 y 27.3. Observe que hay algunas discrepancias entre las aproximaciones. En ambos métodos numéricos, los errores se reducen al disminuir sus respectivos tamaños de paso. Aunque las dos técnicas funcionan bien en el presente caso, se prefiere el procedimiento por diferencias finitas debido a la facilidad con la que se puede extender a casos más complicados.

La condición de frontera fija (o *Dirichlet*) que se usó en el ejemplo anterior es sólo uno de varios tipos que se utilizan comúnmente en ingeniería y ciencia. Una alternativa común, que se llama *condición de frontera Neumann*, es el caso cuando la derivada está dada.

Se puede usar el modelo de barra caliente para demostrar cómo se puede incorporar la condición de frontera derivada en el enfoque de diferencia finita,

$$0 = \frac{d^2T}{dx^2} + h'(T_\infty - T)$$

Sin embargo, en contraste con las discusiones previas, se recomendará una condición de frontera derivada en un extremo de la barra,

Tabla 27.1 Comparación de la solución analítica exacta con los métodos de disparo y de diferencias finitas.

x	Verdadera	Método de disparo	Diferencias finitas
0	40	40	40
2	65.9518	65.9520	65.9698
4	93.7478	93.7481	93.7785
6	124.5036	124.5039	124.5382
8	159.4534	159.4538	159.4795
10	200	200	200

$$\frac{dT}{dx}(0) = T'_a$$

$$T(L) = T_b$$

Así, se tiene una condición de frontera derivada en un extremo del dominio de la solución, y una condición de frontera fija en el otro extremo.

Como se hizo en el ejemplo 27.3, la barra se divide en una serie de nodos y se aplica una versión de diferencias finitas de la ecuación diferencial [ecuación (27.3)] a cada nodo interior. Sin embargo, dado que su temperatura no está especificada, también se debe incluir el nodo del lado izquierdo. Escribiendo la ecuación (27.3) para este nodo se obtiene

$$-T_{-1} + (2 + h'\Delta x^2)T_0 - T_1 = h'\Delta x^2 T_\infty \tag{27.3a}$$

Observe que para esta ecuación se necesita un nodo imaginario (–1) ubicado a la izquierda del extremo de la barra. Aunque este punto exterior podría parecer una dificultad, en realidad sirve como el vehículo para incorporar al problema la condición de frontera derivada. Esto se hace representando la primera derivada en la dimensión x en (0) por la diferencia centrada

$$\frac{dT}{dx} = \frac{T_1 - T_{-1}}{2\Delta x}$$

donde se puede despejar

$$T_{-1} = T_1 - 2\Delta x \frac{dT}{dx}$$

Ahora se tiene una formula para T_{-1} que realmente refleja el impacto de la derivada. Se puede sustituir en la ecuación (27.3*a*) para obtener

$$(2 + h'\Delta x^2)T_0 - 2T_1 = h'\Delta x^2 T_\infty - 2\Delta x \frac{dT}{dx} \tag{27.3b}$$

En consecuencia, se ha incorporado la derivada al balance.

Un ejemplo común de la condición de frontera derivada es el caso en el que el extremo de la barra está aislado. En este caso, la derivada se ajusta a cero. Esta conclusión se deduce directamente de la ley de Fourier, que dice que el flujo de calor es directamente proporcional al gradiente de temperatura. Así, aislar una frontera significa que el flujo de calor (y en consecuencia el gradiente) debe ser cero.

Además de los métodos por diferencias finitas y de disparo, existen otras técnicas para resolver problemas con valores en la frontera. Algunas de éstas se describen en la parte ocho y comprenden soluciones en estado estacionario (capítulo 29) y transitorio (capítulo 30) de problemas con valores en la frontera en dos dimensiones, usando diferencias finitas y soluciones en estado estacionario de problemas unidimensionales con el método del elemento finito (capítulo 31).

27.2 PROBLEMAS DE VALORES PROPIOS

Los problemas de *valores propios*, o característicos o eigenvalores, constituyen una clase especial de problemas con valores en la frontera, que son comunes en el contexto de problemas de ingeniería que implican vibraciones, elasticidad y otros sistemas oscilantes. Además, se utilizan en una amplia variedad de contextos en ingeniería que van más allá de los problemas con valores en la frontera. Antes de describir los métodos numéricos para resolver estos problemas, revisaremos alguna información como antecedente. Ésta comprende el análisis de la importancia tanto matemática como ingenieril de los valores propios.

27.2.1 Antecedentes matemáticos

En la parte tres se estudiaron métodos para resolver sistemas de ecuaciones algebraicas lineales de la forma general

$$[A]\{X\} = \{B\}$$

Tales sistemas se llaman *no homogéneos* debido a la presencia del vector $\{B\}$ en el lado derecho de la igualdad. Si las ecuaciones que constituyen tal sistema son linealmente independientes (es decir, que tienen un determinante distinto de cero), tendrán una solución única. En otras palabras, existe un conjunto de valores x que satisface las ecuaciones.

En cambio, un sistema algebraico lineal homogéneo tiene la forma general:

$$[A]\{X\} = 0$$

Aunque son posibles las soluciones no triviales (es decir, soluciones distintas a todas las $x = 0$) para tales sistemas, generalmente no son únicas. Más bien, las ecuaciones simultáneas establecen relaciones entre las x que se pueden satisfacer con diferentes combinaciones de valores.

Los problemas de valores propios relacionados con la ingeniería tienen la forma general:

$$\begin{array}{rrrr} (a_{11}-\lambda)x_1 + & a_{12}x_2 + \cdots + & a_{1n}x_n & = 0 \\ a_{21}x_1 + & (a_{22}-\lambda)x_2 + \cdots + & a_{2n}x_n & = 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1}x_1 + & a_{n2}x_2 + \cdots + & (a_{nn}-\lambda)x_n & = 0 \end{array}$$

donde λ es un parámetro desconocido llamado *valor propio* o *característico* o *eigenvalor.* Una solución $\{X\}$ de este sistema se le conoce como *vector propio* (vector característico o eigenvector). El conjunto de ecuaciones anterior también se expresa de manera concisa como:

$$\big[[A]-\lambda[I]\big]\{X\} = 0 \tag{27.4}$$

La solución de la ecuación (27.4) depende de la determinación del valor de λ. Una manera de obtenerlo se basa en el hecho de que el determinante de la matriz $[[A]-\lambda[I]]$ debe ser igual a cero para que existan soluciones no triviales. La expansión del determinante será un polinomio en función de λ. Las raíces de este polinomio son los valores propios. En la siguiente sección se presentará un ejemplo de dicho procedimiento.

27.2.2 Antecedentes físicos

El sistema masa-resorte de la figura 27.5*a* es un ejemplo simple para ilustrar cómo se presentan los valores propios en los problemas físicos. También ayudará a entender algunos de los conceptos matemáticos presentados en la sección anterior.

Para simplificar el análisis, suponga que en cada masa no actúan fuerzas externas o de amortiguamiento. Además, considere que cada resorte tiene la misma longitud natural l y la misma constante de elasticidad k. Por último, suponga que el desplazamiento de cada resorte se mide en relación con su sistema coordenado local con un origen en la posición de equilibrio del resorte (figura 27.5*a*). Bajo estas consideraciones, se emplea la segunda ley de Newton para desarrollar un balance de fuerzas para cada masa (recuerde la sección 12.4),

$$m_1\frac{d^2x_1}{dt^2} = -kx_1 + k(x_2 - x_1)$$

y

$$m_2\frac{d^2x_2}{dt^2} = -k(x_2 - x_1) - kx_2$$

donde x_i es el desplazamiento de la masa i respecto de su posición de equilibrio (figura 27.5*b*). Estas ecuaciones se expresan como:

$$m_1\frac{d^2x_1}{dt^2} - k(-2x_1 + x_2) = 0 \tag{27.5a}$$

$$m_2\frac{d^2x_2}{dt^2} - k(x_1 - 2x_2) = 0 \tag{27.5b}$$

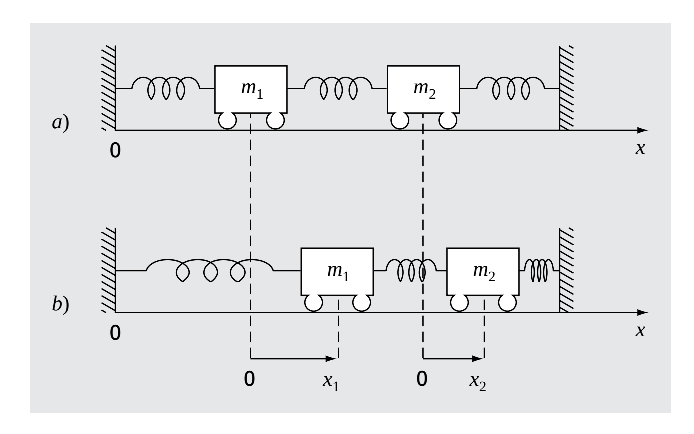


Figura 27.5
Colocando las masas alejadas de su posición de equilibrio se crean fuerzas en los resortes que, después de liberados, hacen oscilar las masas. Las posiciones de las masas se pueden referir a coordenadas locales con orígenes en sus respectivas posiciones de equilibrio.

De la teoría de vibraciones, se conoce que las soluciones de la ecuación (27.5) pueden tomar la forma:

$$x_i = A_i \text{ sen}(\omega t) \tag{27.6}$$

donde A_i = la amplitud de la vibración de la masa i y ω = la frecuencia de la vibración, que es igual a:

$$\omega = \frac{2\pi}{T_p} \tag{27.7}$$

donde T_p es el periodo. De la ecuación (27.6) se tiene que:

$$x_i'' = -A_i\omega^2 \text{sen}\,(\omega t) \tag{27.8}$$

Las ecuaciones (27.6) y (27.8) se sustituyen en las ecuaciones (27.5), y después de agrupar términos, se expresan como:

$$\left(\frac{2k}{m_1} - \omega^2\right)A_1 - \frac{k}{m_1}A_2 = 0 \tag{27.9a}$$

$$-\frac{k}{m_2}A_1 + \left(\frac{2k}{m_2} - \omega^2\right)A_2 = 0 \tag{27.9b}$$

Una comparación entre las ecuaciones (27.9) y (27.4) indica que ahora la solución se redujo a un problema de valores propios.

EJEMPLO 27.4 Valores propios y vectores propios para un sistema masa-resorte

Planteamiento del problema Evalúe los valores propios y los vectores propios de la ecuación (27.9) en el caso donde $m_1 = m_2 = 40$ kg y $k = 200$ N/m.

Solución Sustituyendo los valores de los parámetros en las ecuaciones (27.9) se obtiene:

$$(10 - \omega^2)A_1 - 5A_2 = 0$$

$$-5A_1 + (10 - \omega^2)A_2 = 0$$

El determinante de este sistema es [recuerde la ecuación (9.3)]:

$$(\omega^2)^2 - 20\omega^2 + 75 = 0$$

que puede resolverse con la fórmula cuadrática para $\omega^2 = 15$ y 5 s^{-2}. Por lo tanto, las frecuencias de las vibraciones de las masas son $\omega = 3.873$ s^{-1} y 2.236 s^{-1}, respectivamente. Estos valores se utilizan para determinar los periodos de las vibraciones con la ecuación (27.7). Para el primer periodo, T_p = 1.62 s; y para el segundo, T_p = 2.81 s.

Como se estableció en la sección 27.2.1, no es posible obtener un conjunto único de valores para las incógnitas. Sin embargo, se pueden especificar relaciones entre éstas sustituyendo los valores propios en las ecuaciones. Por ejemplo, para el primero ($\omega^2 = 15$ s^{-2}), $A_1 = -A_2$. Para el segundo ($\omega^2 = 5$ s^{-2}), $A_1 = A_2$.

Este ejemplo proporciona información valiosa con respecto al comportamiento del sistema de la figura 27.5. Además de su periodo, sabemos que si el sistema está vibrando en el primer modo, la amplitud de la segunda masa será igual, pero de signo opuesto a la amplitud de la primera. Como se observa en la figura 27.6*a*, las masas vibran alejándose, y después acercándose de manera indefinida.

En el segundo modo, las dos masas tienen igual amplitud todo el tiempo. Así, como se observa en la figura 27.6*b*, vibran hacia atrás y hacia adelante sincronizadas. Deberá observarse que la configuración de las amplitudes ofrece una guía para ajustar sus valores iniciales para alcanzar un movimiento puro en cualquiera de los dos modos. Cualquier otra configuración llevará a la superposición de los modos de vibración (recuerde el capítulo 19).

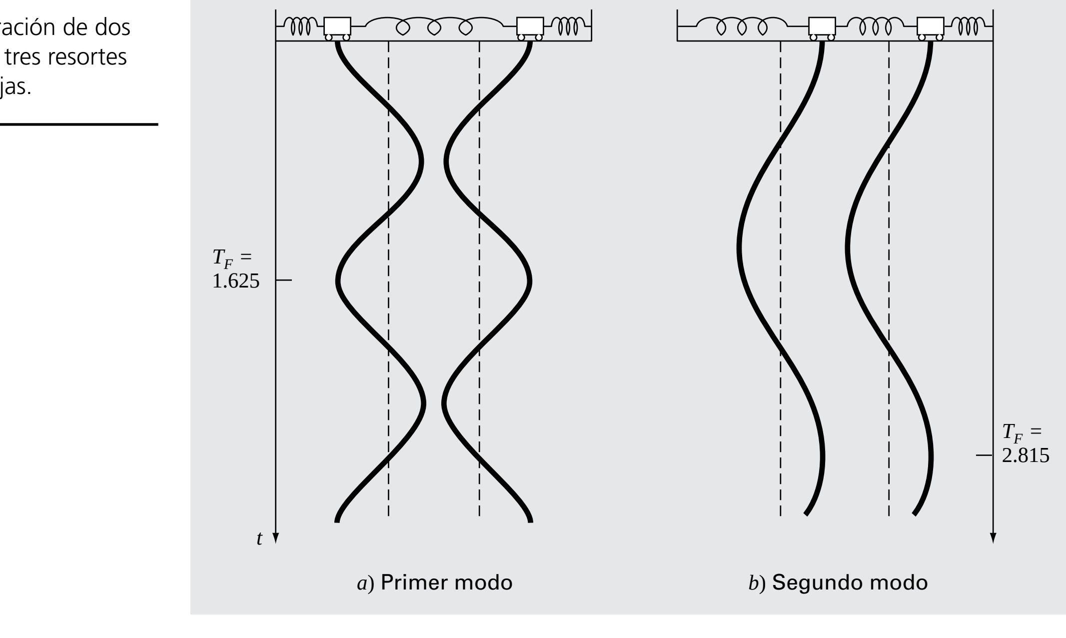


Figura 27.6
Principales modos de vibración de dos masas iguales unidas por tres resortes idénticos entre paredes fijas.

27.2.3 Un problema de valores en la frontera

Ahora que hemos estudiado los valores propios, volvemos al tipo de problemas que es el objeto de este capítulo: problemas con valores en la frontera para ecuaciones diferenciales ordinarias. La figura 27.7 muestra un sistema físico que puede servir como un ejemplo para examinar este tipo de problemas.

La curvatura de una columna delgada sujeta a una carga axial P se modela mediante

$$\frac{d^2y}{dx^2} = \frac{M}{EI} \tag{27.10}$$

donde d^2y/dx^2 especifica la curvatura, M = momento de flexión, E = módulo de elasticidad e I = momento de inercia de la sección transversal con respecto a su eje. Considerando el diagrama de cuerpo libre de la figura 27.7*b*, es claro que el momento de flexión en x es $M = -Py$. Sustituyendo este valor en la ecuación (27.10) se obtiene:

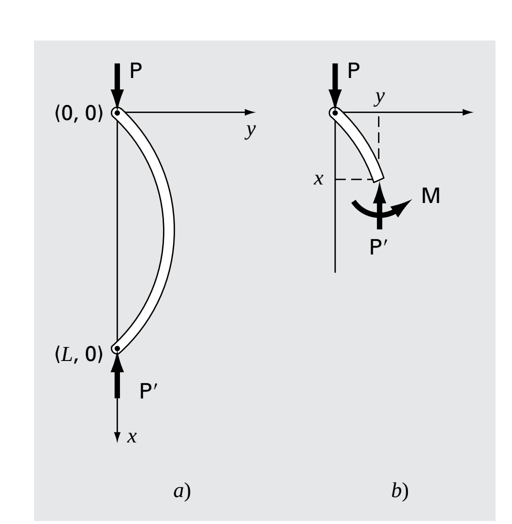


Figura 27.7
a) Barra delgada. *b*) Diagrama de cuerpo libre de la barra.

$$\frac{d^2y}{dx^2} + p^2y = 0 \tag{27.11}$$

donde

$$p^2 = \frac{P}{EI} \tag{27.12}$$

Para el sistema de la figura 27.7, sujeto a las condiciones de frontera

$$y(0) = 0 \tag{27.13a}$$

$$y(L) = 0 \tag{27.13b}$$

la solución general de la ecuación (27.11) es:

$$y = A\,\text{sen}(px) + B\cos(px) \tag{27.14}$$

donde A y B son las constantes esenciales y arbitrarias de la integración que serán evaluadas por medio de las condiciones de frontera. De acuerdo con la primera condición [ecuación (27.13*a*)],

$$0 = A\,\text{sen}(0) + B\cos(0)$$

Por lo tanto, concluimos que $B = 0$.

Se tiene, de acuerdo con la segunda condición [ecuación (27.13*b*)],

$$0 = A \operatorname{sen}(pL) + B \cos(pL)$$

Pero, puesto que $B = 0$, entonces, $A \operatorname{sen}(pL) = 0$. Como $A = 0$ representa una solución trivial, concluimos que $\operatorname{sen}(pL) = 0$. Para que esta igualdad se cumpla,

$$pL = n\pi \qquad \text{para } n = 1, 2, 3, \ldots\ \pi \tag{27.15}$$

Así, existe un número infinito de valores que satisfacen las condiciones de frontera. De la ecuación (27.15) se despeja

$$p = \frac{n\pi}{L} \qquad \text{para } n = 1, 2, 3, \ldots \tag{27.16}$$

los cuales son los valores propios para la columna.

La figura 27.8, que muestra la solución para los primeros cuatro valores propios, puede ofrecer una ilustración sobre el significado físico de los resultados. Cada valor propio corresponde a una manera en la que la columna se dobla o pandea. Combinando las ecuaciones (27.12) y (27.16), se obtiene:

$$P = \frac{n^2\pi^2 EI}{L^2} \qquad \text{para } n = 1, 2, 3, \ldots \tag{27.17}$$

Éstas se consideran *cargas de pandeo*, pues representan los niveles a los cuales se mueve la columna a cada configuración de pandeo sucesivo. En un sentido práctico, usualmente el primer valor es el de interés, ya que, en general, las fallas ocurren cuando la columna se pandea primero. Así, una carga crítica se define como:

$$P = \frac{\pi^2 EI}{L^2}$$

que se conoce como *fórmula de Euler.*

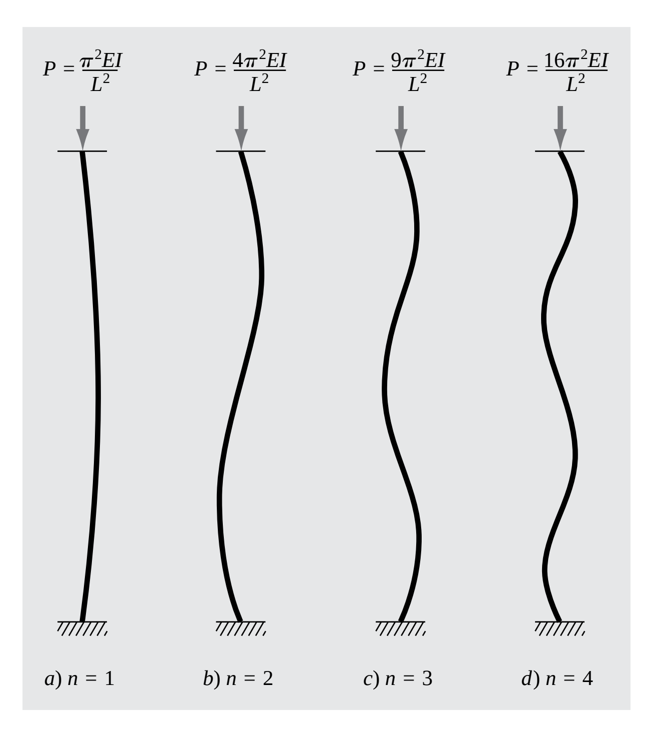


Figura 27.8
Los primeros cuatro valores propios de la barra delgada de la figura 27.7.

EJEMPLO 27.5 Análisis de valores propios de una columna cargada axialmente

Planteamiento del problema Una columna de madera cargada axialmente tiene las siguientes características: $E = 10 \times 10^9$ Pa, $I = 1.25 \times 10^{-5}$ m^4 y $L = 3$ m. Determine los primeros ocho valores propios y las correspondientes cargas de pandeo.

Solución Se utilizan las ecuaciones (27.16) y (27.17) para calcular

n	*p*, m^{-2}	*P*, kN
1	1.0472	137.078
2	2.0944	548.311
3	3.1416	1 233.701
4	4.1888	2 193.245
5	5.2360	3 426.946
6	6.2832	4 934.802
7	7.3304	6 716.814
8	8.3776	8 772.982

La carga crítica de pandeo es, por lo tanto, 137.078 kN.

Aunque las soluciones analíticas del tipo antes obtenido son útiles, a menudo es difícil o imposible obtenerlas. Normalmente esto ocurre cuando se trata con sistemas complicados o con aquellos que tienen propiedades heterogéneas. En tales casos, los métodos numéricos del tipo que a continuación se describirán son la única alternativa práctica.

27.2.4 El método del polinomio

La ecuación (27.11) se puede resolver numéricamente sustituyendo la segunda derivada por una aproximación en diferencias divididas finitas centradas (figura 23.3), lo que da

$$\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} + p^2 y_i = 0$$

la cual se expresa como:

$$y_{i-1} - (2 - h^2p^2)y_i + y_{i+1} = 0 \tag{27.18}$$

Al escribir esta ecuación para una serie de nodos a lo largo del eje de la columna, se obtiene un sistema de ecuaciones homogéneo. Por ejemplo, si la columna se divide en cinco segmentos (es decir, cuatro nodos interiores), el resultado es:

$$\begin{bmatrix} (2-h^2p^2) & -1 & 0 & 0 \\ -1 & (2-h^2p^2) & -1 & 0 \\ 0 & -1 & (2-h^2p^2) & -1 \\ 0 & 0 & -1 & (2-h^2p^2) \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{Bmatrix} = 0 \tag{27.19}$$

La expansión del determinante del sistema da un polinomio, cuyas raíces son los valores propios. Este procedimiento, llamado el *método del polinomio*, se aplica en el siguiente ejemplo.

EJEMPLO 27.6 El método del polinomio

Planteamiento del problema Emplee el método del polinomio para determinar los valores propios de la columna cargada axialmente del ejemplo 27.5, usando *a*) uno, *b*) dos, *c*) tres y *d*) cuatro nodos interiores.

Solución

a) Al escribir la ecuación (27.18) para un nodo interior, se obtiene (h = 3/2)

$$-(2 - 2.25p^2)y_1 = 0$$

Así, en este caso sencillo, el valor propio se analiza igualando el determinante con cero

$$2 - 2.25p^2 = 0$$

obteniendo $p = \pm 0.9428$, que es aproximadamente 10% menor que el valor exacto de 1.0472 obtenido en el ejemplo 27.4.

b) Para dos nodos interiores (h = 3/3), la ecuación (27.18) se escribe como

$$\begin{bmatrix} (2-p^2) & -1 \\ -1 & (2-p^2) \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \end{Bmatrix} = 0$$

La expansión del determinante da

$$(2 - p^2)^2 - 1 = 0$$

de donde se obtiene $p = \pm 1$ y ± 1.73205. De esta manera, el primer valor propio ahora es aproximadamente 4.5% menor, y se obtiene un segundo valor propio que es aproximadamente 17% menor.

c) Para tres puntos interiores (h = 3/4), la ecuación (27.18) se escribe como:

$$\begin{bmatrix} 2-0.5625p^2 & -1 & 0 \\ -1 & 2-0.5625p^2 & -1 \\ 0 & -1 & 2-0.5625p^2 \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \\ y_3 \end{Bmatrix} = 0 \tag{E27.6.1}$$

El determinante se iguala a cero y se expande para dar:

$$(2 - 0.5625p^2)^3 - 2(2 - 0.5625p^2) = 0$$

Para que esta ecuación se satisfaga, $2 - 0.5625p^2 = 0$ y $2 - 0.5625p^2 = \sqrt{2}$. Por lo tanto, los primeros tres valores propios se determinan como:

$p = \pm 1.0205 \qquad |\varepsilon_t| = 2.5\%$
$p = \pm 1.8856 \qquad |\varepsilon_t| = 10\%$
$p = \pm 2.4637 \qquad |\varepsilon_t| = 22\%$

d) Para cuatro puntos interiores ($h = 3/5$), el resultado es la ecuación (27.19) con $2 - 0.36p^2$ sobre la diagonal. Igualando a cero el determinante y expandiéndolo, se tiene:

$$(2 - 0.36p^2)^4 - 3(2 - 0.36p^2)^2 + 1 = 0$$

que se resuelve para los primeros cuatro valores propios

$p = \pm 1.0301 \qquad |\varepsilon_t| = 1.6\%$
$p = \pm 1.9593 \qquad |\varepsilon_t| = 6.5\%$
$p = \pm 2.6967 \qquad |\varepsilon_t| = 14\%$
$p = \pm 3.1702 \qquad |\varepsilon_t| = 24\%$

La tabla 27.2, que resume los resultados de este ejemplo, ilustra algunos aspectos fundamentales del método polinomial. Conforme la segmentación se refina más, se determinan valores propios adicionales, y los valores previamente determinados se vuelven progresivamente más exactos. Así, el procedimiento es muy adecuado para los casos donde se requieren los valores propios.

Tabla 27.2 Los resultados de aplicar el método del polinomio a una columna cargada axialmente. Los números entre paréntesis representan el valor absoluto del error relativo porcentual verdadero.

		Método del polinomio			
Valor propio	**Verdadero**	***h* = 3/2**	***h* = 3/3**	***h* = 3/4**	***h* = 3/5**
1	1.0472	0.9428 (10%)	1.0000 (4.5%)	1.0205 (2.5%)	1.0301 (1.6%)
2	2.0944		1.7321 (21%)	1.8856 (10%)	1.9593 (65%)
3	3.1416			2.4637 (22%)	2.6967 (14%)
4	4.1888				3.1702 (24%)

27.2.5 El método de potencias

El *método de potencias* es un procedimiento iterativo que sirve para determinar el valor propio mayor. Con ligeras modificaciones, también puede ser útil para determinar los valores menor e intermedio. Como ventaja adicional, el vector propio correspondiente se obtiene como parte del método.

Determinación del valor propio mayor Para implementar el método de potencias, el sistema que se analiza debe expresarse en la forma:

$$[A]\{X\} = \lambda \{X\} \tag{27.20}$$

Como se ilustra en el siguiente ejemplo, la ecuación (27.20) es la base para una técnica de solución iterativa que, finalmente, proporciona el valor propio mayor y su vector propio asociado.

EJEMPLO 27.7 Método de potencias para el valor propio mayor

Planteamiento del problema Con el método de potencias determine el valor propio mayor para el inciso *c*) del ejemplo 27.6.

Solución Primero, el sistema se escribe en la forma de la ecuación (27.20),

$$\begin{aligned} 3.556x_1 - 1.778x_2 \qquad\qquad &= \lambda x_1 \\ -1.778x_1 + 3.556x_2 - 1.778x_3 &= \lambda x_2 \\ -1.778x_2 + 3.556x_3 &= \lambda x_3 \end{aligned}$$

Después, suponiendo que las x del lado izquierdo de la ecuación son iguales a 1,

$$\begin{aligned} 3.556(1) - 1.778(1) \qquad\qquad &= 1.778 \\ -1.778(1) + 3.556(1) - 1.778(1) &= 0 \\ -1.778(1) + 3.556(1) &= 1.778 \end{aligned}$$

Luego, el lado derecho se normaliza con 1.778 para hacer que el elemento mayor sea igual a:

$$\begin{Bmatrix} 1.778 \\ 0 \\ 1.778 \end{Bmatrix} = 1.778 \begin{Bmatrix} 1 \\ 0 \\ 1 \end{Bmatrix}$$

Así, la primera estimación del valor propio es 1.778. Esta iteración se expresa en forma matricial como:

$$\begin{bmatrix} 3.556 & -1.778 & 0 \\ -1.778 & 3.556 & -1.778 \\ 0 & -1.778 & 3.556 \end{bmatrix} \begin{Bmatrix} 1 \\ 1 \\ 1 \end{Bmatrix} = \begin{Bmatrix} 1.778 \\ 0 \\ 1.778 \end{Bmatrix} = 1.778 \begin{Bmatrix} 1 \\ 0 \\ 1 \end{Bmatrix}$$

La siguiente iteración consiste en multiplicar $[A]$ por $\lfloor 1 \quad 0 \quad 1 \rfloor^T$ para dar:

$$\begin{bmatrix} 3.556 & -1.778 & 0 \\ -1.778 & 3.556 & -1.778 \\ 0 & -1.778 & 3.556 \end{bmatrix} \begin{Bmatrix} 1 \\ 0 \\ 1 \end{Bmatrix} = \begin{Bmatrix} 3.556 \\ -3.556 \\ 3.556 \end{Bmatrix} = 3.556 \begin{Bmatrix} 1 \\ -1 \\ 1 \end{Bmatrix}$$

Por lo tanto, el valor propio estimado en la segunda iteración es 3.556, que puede emplearse para determinar el error estimado

$$|\varepsilon_a| = \left| \frac{3.556 - 1.778}{3.556} \right| 100\% = 50\%$$

Luego, el proceso puede repetirse.

Tercera iteración:

$$\begin{bmatrix} 3.556 & -1.778 & 0 \\ -1.778 & 3.556 & -1.778 \\ 0 & -1.778 & 3.556 \end{bmatrix} \begin{Bmatrix} 1 \\ -1 \\ 1 \end{Bmatrix} = \begin{Bmatrix} 5.334 \\ -7.112 \\ 5.334 \end{Bmatrix} = -7.112 \begin{Bmatrix} -0.75 \\ 1 \\ -0.75 \end{Bmatrix}$$

donde $|\varepsilon_a|$ = 150% (que es alto debido al cambio de signo).

Cuarta iteración:

$$\begin{bmatrix} 3.556 & -1.778 & 0 \\ -1.778 & 3.556 & -1.778 \\ 0 & -1.778 & 3.556 \end{bmatrix} \begin{Bmatrix} -0.75 \\ 1 \\ -0.75 \end{Bmatrix} = \begin{Bmatrix} -4.445 \\ 6.223 \\ -4.445 \end{Bmatrix} = 6.223 \begin{Bmatrix} -0.714 \\ 1 \\ -0.714 \end{Bmatrix}$$

donde $|\varepsilon_a|$ = 214% (de nuevo, muy alto debido al cambio de signo).

Quinta iteración:

$$\begin{bmatrix} 3.556 & -1.778 & 0 \\ -1.778 & 3.556 & -1.778 \\ 0 & -1.778 & 3.556 \end{bmatrix} \begin{Bmatrix} -0.714 \\ 1 \\ -0.714 \end{Bmatrix} = \begin{Bmatrix} -4.317 \\ 6.095 \\ -4.317 \end{Bmatrix} = 6.095 \begin{Bmatrix} -0.708 \\ 1 \\ -0.708 \end{Bmatrix}$$

Así, el factor normalizado converge al valor de 6.070 (= 2.4637^2) obtenido en el inciso *c*) del ejemplo 27.6.

Tenga en cuenta que en algunas ocasiones el método de potencias convergerá al segundo valor propio más grande, en lugar de hacerlo al primero. James, Smith y Wolford (1985) presentan un caso así. Otros casos especiales se analizan en Fadeev y Fadeeva (1963).

Determinación del valor propio menor. En ingeniería existen problemas donde nos interesa determinar el valor propio menor. Tal es el caso de la barra en la figura 27.7, donde el valor propio menor se utilizó para identificar una carga de pandeo crítica. Esto puede realizarse aplicando el método de potencias a la matriz inversa de [*A*]. En este caso, el método de potencias converge al valor mayor de $1/\lambda$ (en otras palabras, el valor menor de λ).

EJEMPLO 27.8 Método de potencias para el valor propio menor

Planteamiento del problema Emplee el método de potencias para determinar el valor propio menor en el inciso *c*) del ejemplo 27.6.

Solución Después de dividir la ecuación (E27.6.1) entre h^2 (= 0.5625), se evalúa su matriz inversa como:

$$[A]^{-1} = \begin{bmatrix} 0.422 & 0.281 & 0.141 \\ 0.281 & 0.562 & 0.281 \\ 0.141 & 0.281 & 0.422 \end{bmatrix}$$

Usando el mismo formato del ejemplo 27.7, el método de potencias se aplica a esta matriz.

Primera iteración:

$$\begin{bmatrix} 0.422 & 0.281 & 0.141 \\ 0.281 & 0.562 & 0.281 \\ 0.141 & 0.281 & 0.422 \end{bmatrix} \begin{Bmatrix} 1 \\ 1 \\ 1 \end{Bmatrix} = \begin{Bmatrix} 0.884 \\ 1.124 \\ 0.884 \end{Bmatrix} = 1.124 \begin{Bmatrix} 0.751 \\ 1 \\ 0.751 \end{Bmatrix}$$

Segunda iteración:

$$\begin{bmatrix} 0.422 & 0.281 & 0.141 \\ 0.281 & 0.562 & 0.281 \\ 0.141 & 0.281 & 0.422 \end{bmatrix} \begin{Bmatrix} 0.751 \\ 1 \\ 0.751 \end{Bmatrix} = \begin{Bmatrix} 0.704 \\ 0.984 \\ 0.704 \end{Bmatrix} = 0.984 \begin{Bmatrix} 0.715 \\ 1 \\ 0.715 \end{Bmatrix}$$

donde $|\varepsilon_a|$ = 14.6%.

Tercera iteración:

$$\begin{bmatrix} 0.422 & 0.281 & 0.141 \\ 0.281 & 0.562 & 0.281 \\ 0.141 & 0.281 & 0.422 \end{bmatrix} \begin{Bmatrix} 0.715 \\ 1 \\ 0.715 \end{Bmatrix} = \begin{Bmatrix} 0.684 \\ 0.964 \\ 0.684 \end{Bmatrix} = 0.964 \begin{Bmatrix} 0.709 \\ 1 \\ 0.709 \end{Bmatrix}$$

donde $|\varepsilon_a| = 4\%$.

Así, después de sólo tres iteraciones, el resultado converge al valor de 0.9602, que es el recíproco del valor propio menor, 1.0205 ($= \sqrt{1/0.9602}$), obtenido en el ejemplo 27.6*c*.

Determinación de valores propios intermedios Después de encontrar el mayor de los valores propios, es posible determinar los siguientes más grandes reemplazando la matriz original por una que incluya sólo los valores propios restantes. El proceso de eliminar el valor propio mayor conocido se llama *deflación*. La técnica explicada aquí, el *método de Hotelling,* está diseñada para matrices simétricas. Esto es porque aprovecha la ortogonalidad de los vectores propios de tales matrices, los cuales se expresan como:

$$\{X\}_i^T\{X\}_j = \begin{cases} 0 & \text{para } i \neq j \\ 1 & \text{para } i = j \end{cases} \tag{27.21}$$

donde los componentes del vector propio $\{X\}$ se han normalizado de forma tal que $\{X\}^T\{X\} = 1$; es decir, que la suma de los cuadrados de los componentes sea igual a 1. Esto se puede llevar a cabo dividiendo cada uno de los elementos entre el factor normalizado

$$\sqrt{\sum_{k=1}^{n} x_k^2}$$

Ahora, se calcula una nueva matriz $[A]_2$ como:

$$[A]_2 = [A]_1 - \lambda_1\{X\}_1\{X\}_1^T \tag{27.22}$$

donde $[A]_1$ = la matriz original y λ_1 = el valor propio mayor. Si el método de potencias se aplica a esta matriz, el proceso de iteración converge al segundo valor propio más grande, λ_2. Para demostrarlo, primero multiplicamos por el lado derecho a la ecuación (27.22) por $\{X\}_1$,

$$[A]_2\{X\}_1 = [A]_1\{X\}_1 - \lambda_1\{X\}_1\{X\}_1^T\{X\}_1$$

Considerando el principio de ortogonalidad, esta ecuación se transforma en:

$$[A]_2\{X\}_1 = [A]_1\{X\}_1 - \lambda_1\{X\}_1$$

donde el lado derecho es igual a cero, de acuerdo con la ecuación (27.20). Así, $[A]_2\{X\}_1 = 0$. En consecuencia, $\lambda = 0$ y $\{X\} = \{X\}_1$ es una solución para $[A]_2\{X\} = \lambda\{X\}$. En otras palabras, la matriz $[A]_2$ tiene los valores propios $0, \lambda_2, \lambda_3, \ldots, \lambda_n$. El valor propio mayor λ_1 se reemplazó con un 0 y, por lo tanto, el método de potencias convergerá al siguiente λ_2 más grande.

El proceso anterior puede repetirse generando una nueva matriz $[A]_3$, etc. Aunque, en teoría, este proceso podría continuar para determinar los valores propios restantes, está limitado por el hecho de que en cada paso se arrastran los errores sobre los vectores propios. Por ello, solamente es útil para determinar algunos de los valores propios más altos. Aunque, de alguna manera, esto es una desventaja, se requiere precisamente esta información en muchos problemas de ingeniería.

27.2.6 Otros métodos

Existe una gran variedad de métodos alternativos para resolver problemas de valores propios. La mayoría se basa en un proceso de dos pasos. El primer paso consiste en transformar la matriz original en una forma más simple (por ejemplo, tridiagonal), que conserve todos los valores propios originales. Después, se usan métodos iterativos para determinar estos valores propios.

Muchos de esos procedimientos están diseñados para tipos especiales de matrices. En particular, varias técnicas se dedican a la solución de sistemas simétricos. Por ejemplo, el *método de Jacobi* transforma una matriz simétrica en una matriz diagonal, al eliminar de forma sistemática los términos que están fuera de la diagonal. Por desgracia, el método requiere un enorme número de operaciones, ya que la eliminación de cada elemento distinto de cero a menudo crea un nuevo valor distinto de cero en un elemento previamente anulado. A pesar de que se requiere muchísimo tiempo para eliminar todos los elementos distintos de cero fuera de la diagonal, finalmente la matriz tenderá hacia una forma diagonal. Así, el procedimiento es iterativo en el sentido de que se repite hasta que los términos que están fuera de la diagonal son "suficientemente" pequeños.

El *método de Given* también implica transformar una matriz simétrica en una forma más simple. No obstante, a diferencia del método de Jacobi, la forma más simple es tridiagonal. Además, difiere en que los ceros creados en posiciones fuera de la diagonal se conservan. En consecuencia, es finito y, por lo tanto, más eficiente que el método de Jacobi.

El *método de Householder* también transforma una matriz simétrica en una forma tridiagonal. Es un método finito más eficiente que el método de Given, debido a que reduce a cero todos los elementos, renglones y columnas que están colocados fuera de la diagonal.

Una vez que se obtiene un sistema tridiagonal mediante el método de Given o de Householder, los pasos restantes buscan hallar los valores propios. Una forma directa para realizar esto es expandir el determinante. El resultado es una secuencia de polinomios que se pueden evaluar iterativamente para los valores propios.

Además de las matrices simétricas, también existen técnicas que están disponibles cuando se requieren todos los valores propios de una matriz general. Éstas incluyen el *método LR* de Rutishauser y el *método QR* de Francis. Aunque este último es menos eficiente, a menudo es el método preferido, ya que es más estable. De hecho, se considera como el mejor método de solución para propósitos generales.

Por último, debemos recordar que las técnicas antes mencionadas de manera común se utilizan conjuntamente para aprovechar sus ventajas respectivas. Por ejemplo, los métodos de Given y de Householder también se aplican a sistemas no simétricos. El resultado no será tridiagonal, sino más bien un tipo especial llamado *forma de Hessenberg.* Un procedimiento es aprovechar la velocidad del método de Householder para transformar la matriz a esta forma y, después, usar el algoritmo estable QR para hallar los valores propios. Información adicional sobre éstos y otros temas relacionados con los valores propios se encuentra en Ralston y Rabinowitz (1978), Wilkinson (1965), Fadeev y Fadeeva (1963) y Householder (1953, 1964). Están disponibles códigos para computadora en diferentes fuentes, como Press y cols. (2007). Rice (1983) analiza los paquetes de software disponibles.

27.3 EDO Y VALORES PROPIOS CON PAQUETES DE SOFTWARE

Los paquetes de software tienen grandes capacidades para resolver EDO y determinar valores propios. En esta sección se explican algunas de las formas en que pueden aplicarse con tal propósito.

27.3.1 Excel

Las capacidades directas de Excel para resolver problemas de valores propios y EDO son limitadas. Sin embargo, si se realiza alguna programación (por ejemplo, macros), se puede combinar con las herramientas de visualización y optimización de Excel para implementar algunas aplicaciones interesantes. En la sección 28.1 se proporciona un ejemplo de cómo se utiliza el Solver de Excel para la estimación de parámetros de una EDO.

27.3.2 MATLAB

Como podría esperarse, el paquete estándar MATLAB tiene excelentes capacidades para determinar valores y vectores propios. Aunque también tiene funciones prediseñadas para resolver EDO. Las soluciones estándar de EDO incluyen dos funciones para implementar el método Runge-Kutta Fehlberg con tamaño de paso adaptativo (recuerde la sección 25.5.2). Éstas son **ODE23**, la cual usa fórmu-

las de segundo y de tercer orden para alcanzar una exactitud media; y **ODE45**, que emplea fórmulas de cuarto y de quinto orden para alcanzar una exactitud alta. El siguiente ejemplo ilustra la manera en que se utilizan para resolver un sistema de EDO.

EJEMPLO 27.9 Uso de MATLAB para valores propios y EDO

Planteamiento del problema Explore cómo se utiliza MATLAB para resolver el siguiente conjunto de EDO no lineales desde $t = 0$ hasta 20:

$$\frac{dx}{dt} = 1.2x - 0.6xy \qquad \frac{dy}{dt} = -0.8y + 0.3xy$$

donde $x = 2$ y $y = 1$ en $t = 0$. Como se verá en el siguiente capítulo (sección 28.2), tales ecuaciones se conocen como *ecuaciones depredador-presa*.

Solución Antes de obtener una solución con MATLAB, usted debe usar un procesador de texto para crear un archivo M que contenga el lado derecho de las EDO. Este archivo M será después utilizado para la solución de la EDO [donde $x = y(1)$ y $y = y(2)$]:

```
function yp = predprey(t,y)
yp = [1.2*y(1)–0.6*y(1)*y(2);–0.8*y(2)+0.3*y(1)*y(2)];
```

Guardamos este archivo M con el nombre: predprey.m.

Después, inicie con MATLAB, e introduzca las siguientes instrucciones para especificar el intervalo de integración y las condiciones iniciales:

```
>> tspan = [0,20];
>> y0=[2,1];
```

Luego se pide la solución mediante

```
>> [t,y]=ode23('predprey',tspan,y0);
```

Esta instrucción resolverá entonces las ecuaciones diferenciales en predprey.m en el intervalo definido por `tspan` usando las condiciones iniciales encontradas en `y0`. Los resultados se despliegan tecleando simplemente

```
>> plot(t,y)
```

con lo cual se obtiene la figura 27.9.

Además, también se puede generar una gráfica del estado fásico; es decir, una gráfica de las variables dependientes, una con respecto de la otra mediante

```
>> plot(y(:,1),y(:,2))
```

con lo que se obtiene la figura 27.10.

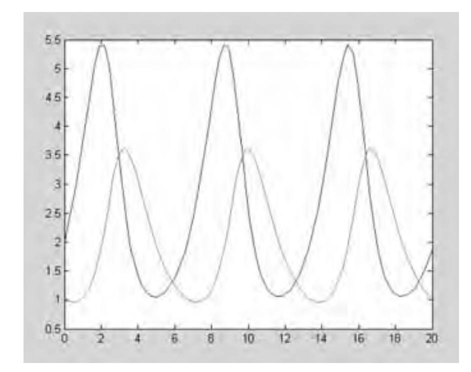


Figura 27.9
Solución del modelo depredador-presa con MATLAB.

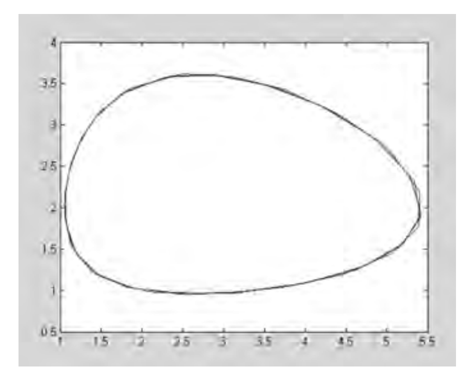


Figura 27.10
Gráfica del estado fásico para el modelo depredador-presa con MATLAB.

En MATLAB también se incluyen funciones diseñadas para sistemas rígidos, definidas por ODE15S y ODE23S. Como se muestra en el siguiente ejemplo, éstas funcionan bien cuando fallan las funciones estándar.

EJEMPLO 27.10 MATLAB para EDO rígidas

Planteamiento del problema La *ecuación de Van der Pol* se puede escribir como:

$$\frac{dy_1}{dt} = y_2$$

$$\frac{dy_2}{dt} = \mu(1 - y_1^2)y_2 - y_1$$

Cuando el parámetro μ es muy grande, el sistema se convierte progresivamente en rígido. Dadas las condiciones iniciales, $y_1(0) = y_2(0) = 1$, use MATLAB para resolver los dos casos siguientes.

a) Para $\mu = 1$, utilice ODE45 para resolver desde $t = 0$ hasta 20.
b) Para $\mu = 1\,000$, utilice ODE23S para resolver desde $t = 0$ hasta 3 000.

Solución

a) Se crea un archivo M para tener las ecuaciones diferenciales

```
function yp = vanderpol(t,y)
yp=[y(2);1*(1-y(1)^2)*y(2)-y(1)];
```

Después, como en el ejemplo 27.9, se llama a ODE45, el resultado se grafica (figura 27.11)

```
>> tspan=[0,20];
>> y0=[1,1];
>> [t,y]=ode45('vanderpol',tspan,y0);
>> plot(t,y(:,1))
```

b) Si se utiliza una solución estándar como ODE45, para el caso rígido ($\mu = 1\,000$), fallará irremediablemente (inténtelo y vea qué sucede); sin embargo, ODE23S hace un trabajo eficiente. Cambie el nuevo valor de μ en el archivo M, la solución se obtiene y se grafica (figura 27.12),

```
>> tspan=[0,3000];
>> y0=[1,1];
>> [t,y]=ode23S('vanderpol',tspan,y0);
>> plot(t,y(:,1))
```

Observe que esta solución tiene bordes más afilados que los del inciso *a*). Ésta es una manifestación visual para la "rigidez" de la solución.

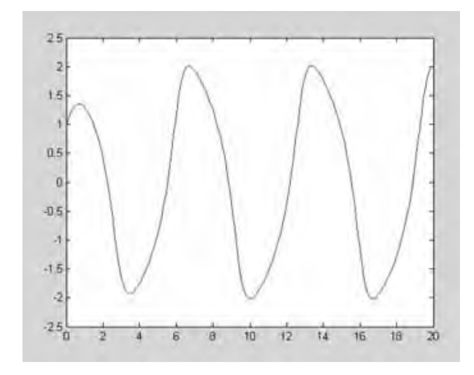


Figura 27.11
La ecuación de Van der Pol en la forma no rígida resuelta con la función ODE45 de MATLAB.

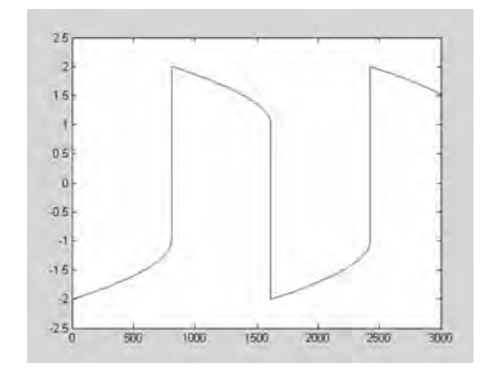


Figura 27.12
La ecuación de Van der Pol en la forma rígida resuelta con la función ODE23S de MATLAB.

Para valores propios, las capacidades también son de una muy fácil aplicación. Recuerde que, en nuestro análisis de sistemas rígidos del capítulo 26, presentamos el sistema rígido definido por la ecuación (26.6). Tales EDO lineales se escriben como un problema de valores propios de la forma

$$\begin{bmatrix} 5-\lambda & -3 \\ -100 & 301-\lambda \end{bmatrix} \begin{Bmatrix} e_1 \\ e_2 \end{Bmatrix} = \{0\}$$

donde λ y $\{e\}$ = el valor propio y vector propio, respectivamente.

MATLAB se puede emplear, entonces, para encontrar tanto los valores propios (d) como los vectores propios (v) con las sencillas instrucciones siguientes:

```
>> a=[5 -3;-100 301];
>> [v,d]=eig(a)

v =
   -0.9477     0.0101
   -0.3191    -0.9999

d =
    3.9899          0
         0   302.0101
```

Así, vemos que los valores propios son muy diferentes en magnitud, lo cual es común en un sistema rígido.

Los valores propios se interpretan reconociendo que la solución general de un sistema de EDO se puede representar como la suma de exponenciales. Por ejemplo, en este caso la solución tendrá la forma:

$$y_1 = c_{11}e^{-3.9899t} + c_{12}e^{-302.0101t}$$

$$y_2 = c_{21}e^{-3.9899t} + c_{22}e^{-302.0101t}$$

donde c_{ij} = la parte de la condición inicial de y_i correspondiente al j-ésimo valor propio. Debe observarse que las c pueden evaluarse a partir de las condiciones iniciales y de los vectores propios. Cualquier buen libro sobre ecuaciones diferenciales, como por ejemplo, el de Boyce y DiPrima (1992), le explicará cómo se puede realizar esto.

Puesto que, en este caso, todos los valores propios son positivos (y, por lo tanto, negativos en la función exponencial), la solución consta de un conjunto de exponenciales en decaimiento. La que tiene el valor propio más grande (en este caso, 302.0101) determinará el tamaño de paso en caso de que utilice una técnica con solución explícita.

27.3.3 Mathcad

Mathcad tiene varias funciones diferentes que resuelven ecuaciones diferenciales y determinan valores propios y vectores propios. La técnica más básica que utiliza Mathcad para resolver sistemas de ecuaciones diferenciales de primer orden es un algoritmo de Runge-Kutta de tamaño fijo de paso, de cuarto orden. Esto lo da la función **rkfixed**. Aunque éste es un buen integrador para todo uso, no siempre es eficiente. Por lo tanto, Mathcad suministra **Rkadapt**, una versión con tamaño variable de paso de **rkfixed**. Es adecuado para funciones que cambian rápidamente en algunas regiones y lentamente en otras. De igual manera, si usted sabe que su solución es una función continua, entonces puede encontrar que la función Mathcad **Bulstoer** trabaja bien. Ésta utiliza el método Bulirsch-Stoer, y a menudo es tanto eficiente como altamente exacto para funciones continuas.

Las ecuaciones diferenciales rígidas constituyen el lado opuesto del espectro. Bajo estas condiciones, la función **rkfixed** puede ser muy ineficiente o inestable. Por lo tanto, Mathcad proporciona dos métodos especiales específicamente diseñados para manejar sistemas rígidos. Estas funciones se llaman **Stiffb** y **Stiffr**, y se basan en el método Bulirsch-Stoer modificado para sistemas rígidos y el método Rosenbrock.

Como ejemplo, se usará Mathcad para resolver las siguientes ecuaciones diferenciales ordinarias (EDO) no lineales:

$$\frac{dy_1}{dt} = 1.2y_1 - 0.6y_1y_2$$
$$\frac{dy_2}{dt} = -0.8y_2 + 0.3y_1y_2$$

con las condiciones iniciales $y_1 = 2$ y $y_2 = 1$. Este sistema, llamado *ecuaciones Lotka-Volterra*, lo usan los ingenieros ambientalistas y los ecólogos para evaluar las interacciones de los depredadores (y_2) y su presa (y_1).

Como en la figura 27.13, primero se usa el símbolo de definición para precisar el vector D(u, y) que contiene los lados derechos de las EDO para su introducción a **rkfixed**. Observe que y_1 y y_2 en las EDO se cambian a y_0 y y_1 para cumplir los requisitos de Mathcad. Además, se definen las condiciones iniciales (y_0), el límite de integración (tf) y el número de valores que se quieren generar (npts). Las soluciones para **rkfixed** con 200 pasos entre $t = 0$ y tf se almacenan en la matriz ysol. La solución se muestra esquemáticamente en la gráfica de la figura 27.13.

En seguida se puede ilustrar cómo evalúa Mathcad los valores propios y los vectores propios. La función **eigenvals(M)** devuelve los valores propios de la matriz cuadrada **M**. La función **eigenvecs(M)** devuelve una matriz que contiene los vectores propios normalizados correspondientes a los vectores propios de **M**, mientras que **eigenvec(M,e)** devuelve el vector propio correspondiente al valor propio **e**. Se pueden ilustrar estas funciones para el sistema dado por [recuerde la ecuación (26.6)]

$$\frac{dy_1}{dt} = -5y_1 + 3y_2$$

$$\frac{dy_2}{dt} = 100y_1 - 301y_2$$

Los resultados se muestran en la figura 27.14. Como los valores propios (aa) son de diferentes magnitudes, el sistema es rígido. Observe que bb contiene el vector propio específico asociado con el valor propio más pequeño. El resultado, cc, es una matriz que contiene ambos vectores propios como sus columnas.

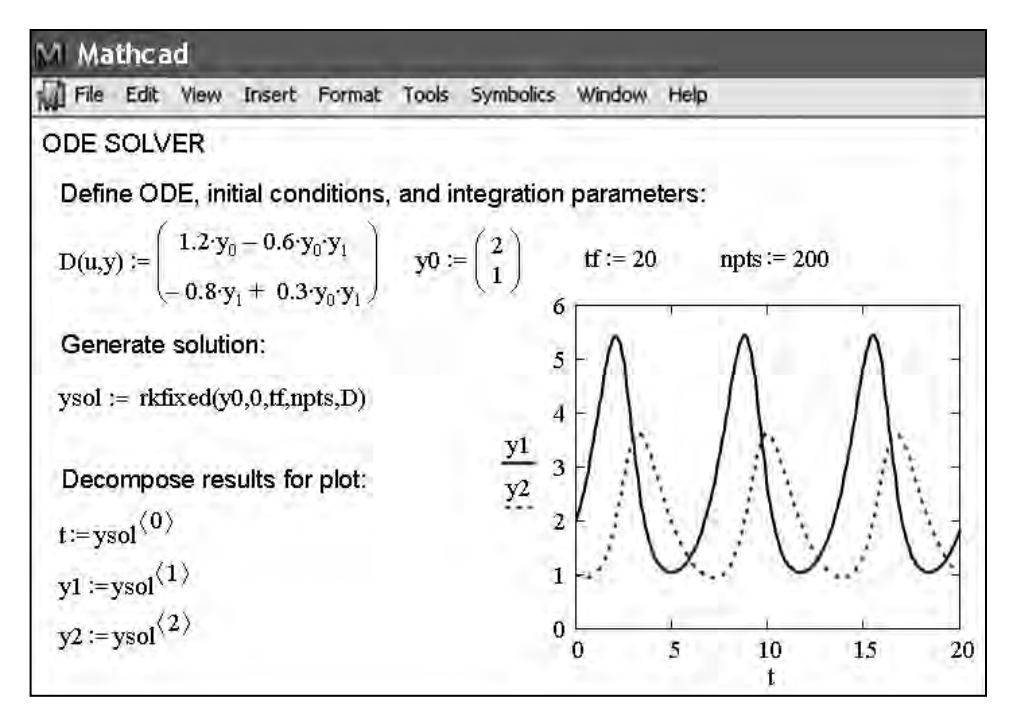


Figura 27.13
Pantalla Mathcad para resolver un sistema de EDO.

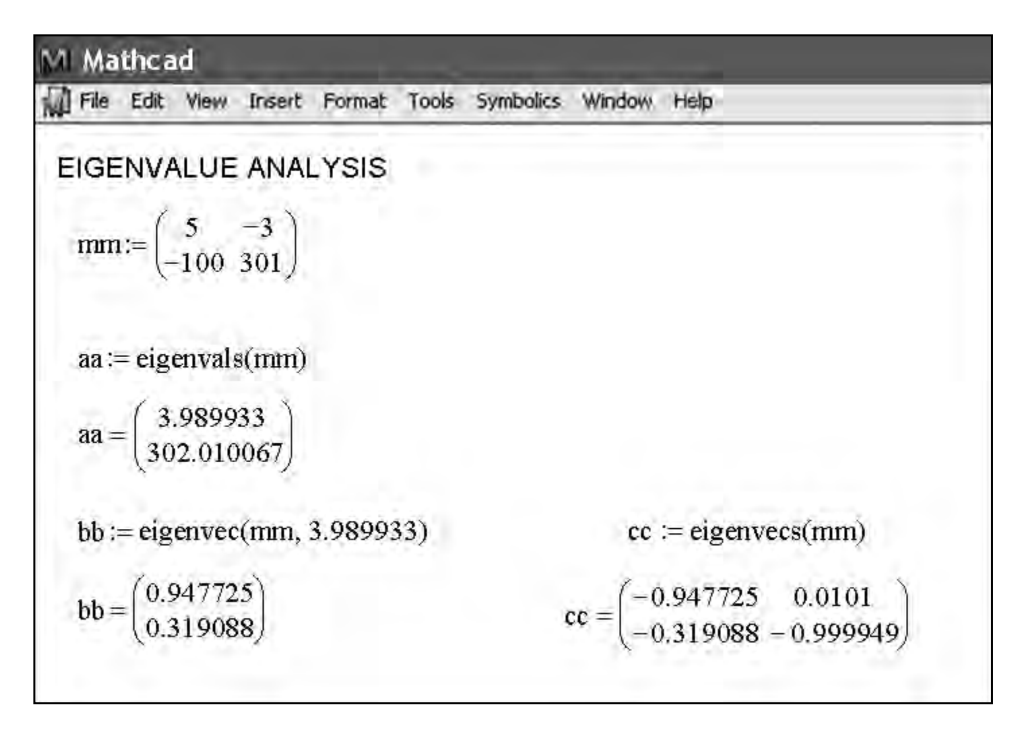


Figura 27.14
Pantalla Mathcad para despejar los valores propios de un sistema de EDO.

PROBLEMAS

27.1 El balance de calor de estado estacionario de una barra se representa como:

$$\frac{d^2T}{dx^2} - 0.15T = 0$$

Obtenga una solución analítica para una barra de 10 m con $T(0) = 240$ y $T(10) = 150$.

27.2 Use el método del disparo para resolver el problema 27.1.

27.3 Use el enfoque de diferencias finitas con $\Delta x = 1$ para resolver el problema 27.1.

27.4 Emplee el método del disparo para resolver

$$7\frac{d^2y}{dx^2} - 2\frac{dy}{dx} - y + x = 0$$

con las condiciones de frontera $y(0) = 5$ y $y(20) = 8$.

27.5 Resuelva el problema 27.4 con el enfoque de diferencias finitas con $\Delta x = 2$.

27.6 Utilice el método del disparo para solucionar

$$\frac{d^2T}{dx^2} - 1\times 10^{-7}(T+273)^4 + 4(150 - T) = 0 \qquad \text{(P27.6.1)}$$

Obtenga una solución para las condiciones de frontera: $T(0) = 200$ y $T(0.5) = 100$.

27.7 Es frecuente que las ecuaciones diferenciales como la que se resolvió en el problema 27.6 se puedan simplificar si se linealizan los términos no lineales. Por ejemplo, para linealizar el término a la cuarta potencia de la ecuación (P27.6.1), se puede usar una expansión en series de Taylor de primer orden, así:

$$1 \times 10^{-7}(T + 273)^4 = 1 \times 10^{-7}(T_b + 273)^4 + 4 \times 10^{-7}(T_b + 273)^3(T - T_b)$$

donde T_b es la temperatura base acerca de la que se linealiza el término. Sustituya esta relación en la ecuación (P27.6.1) y luego resuelva la ecuación lineal resultante con el enfoque de diferencias finitas. Emplee $T_b = 150$ y $\Delta x = 0.01$ para obtener su solución.

27.8 Repita el ejemplo 27.4 pero para tres masas. Elabore una gráfica como la de la figura 27.6 para identificar los principales modos de vibración. Cambie todas las k a 240.

27.9 Vuelva a hacer el ejemplo 27.6, pero para cinco puntos interiores ($h = 3/6$).

27.10 Use menores para expandir el determinante de:

$$\begin{bmatrix} 2-\lambda & 8 & 10 \\ 8 & 4-\lambda & 5 \\ 10 & 5 & 7-\lambda \end{bmatrix}$$

27.11 Emplee el método de potencias para determinar el valor propio más alto y el vector propio correspondiente, para el problema 27.10.

27.12 Emplee el método de potencias para determinar el valor propio más bajo y el vector propio correspondiente para el problema 27.10.

27.13 Desarrolle un programa de cómputo amigable para el usuario a fin de implantar el método del disparo para una EDO lineal de segundo orden. Pruebe el programa con la duplicación del ejemplo 27.1.

27.14 Utilice el programa que desarrolló en el problema 27.13 para resolver los problemas 27.2 y 27.4.
27.15 Desarrolle un programa de computadora amigable para el usuario para implantar la aproximación de diferencias finitas para resolver una EDO lineal de segundo orden. Pruébelo con la duplicación del ejemplo 27.3.
27.16 Utilice el programa desarrollado en el problema 27.15 para resolver los problemas 27.3 y 27.5.
27.17 Desarrolle un programa amigable para el usuario para encontrar el valor propio más alto con el método de la potencia. Pruébelo con la duplicación del ejemplo 27.7.
27.18 Desarrolle un programa amigable para el usuario a fin de resolver el valor propio más pequeño con el método de la potencia. Pruébelo con la duplicación del ejemplo 27.8.
27.19 Emplee la herramienta Solver de Excel para solucionar directamente (es decir, sin linealización) el problema 27.6 con el uso del enfoque de diferencias finitas. Emplee $\Delta x = 0.1$ para obtener su solución.
27.20 Use MATLAB para integrar el siguiente par de EDO, de $t = 0$ a 100:

$$\frac{dy_1}{dt} = 0.35y_1 - 1.6y_1y_2 \qquad \frac{dy_2}{dt} = 0.04y_1y_2 - 0.15y_2$$

donde $y_1 = 1$ y $y_2 = 0.05$ en $t = 0$. Desarrolle una gráfica de espacio fásico (es una gráfica de y_1 *versus* y_2 para t) de sus resultados.
27.21 La ecuación diferencial que sigue se utilizó en la sección 8.4 para analizar la vibración del amortiguador de un auto:

$$1.25\times10^6\frac{d^2x}{dt^2} + 1\times10^7\frac{dx}{dt} + 1.5\times10^9 x = 0$$

Transforme esta ecuación en un par de EDO. *a*) Use MATLAB para resolver las ecuaciones, de $t = 0$ a 0.4, para el caso en que $x = 0.5$, y $dx/dt = 0$ en $t = 0$. *b*) Emplee MATLAB para determinar los valores y vectores propios para el sistema.
27.22 Use MATLAB o Mathcad para integrar:

$$\frac{dx}{dt} = -\sigma x + \sigma y$$

$$\frac{dy}{dt} = rx - y - xz$$

$$\frac{dz}{dt} = -bz + xy$$

donde $\sigma = 10$, $b = 2.666667$ y $r = 28$. Use condiciones iniciales de $x = y = z = 5$, e integre desde $t = 0$ hasta 20.
27.23 Utilice diferencias finitas para resolver la ecuación diferencial ordinaria con valores en la frontera

$$\frac{d^2u}{dx^2} + 6\frac{du}{dx} - u = 2$$

con condiciones de frontera $u(0) = 10$ y $u(2) = 1$. Grafique los resultados de u *versus* x. Utilice $\Delta x = 0.1$.
27.24 Resuelva para la EDO no dimensionada, por medio del método de diferencias finitas, que describa la distribución de la temperatura en una barra circular con fuente interna de calor S.

$$\frac{d^2T}{dr^2} + \frac{1}{r}\frac{dT}{dr} + S = 0$$

en el rango $0 \le r \le 1$, con las condiciones de frontera

$$T(r=1) = 1 \qquad \left.\frac{dT}{dr}\right|_{r=0} = 0$$

para $S = 1$, 10 y 20 K/m^2. Grafique la temperatura *versus* el radio.
27.25 Obtenga el conjunto de ecuaciones diferenciales para un sistema de cuatro resortes y tres masas (figura P27.25) que describa su movimiento en el tiempo. Escriba las tres ecuaciones diferenciales en forma matricial.

[vector de aceleración] + [matriz k/m]

[vector de desplazamiento x] = 0

Observe que cada ecuación ha sido dividida entre la masa. Resuelva para los valores propios y frecuencias naturales para los valores siguientes de masa y constantes de elasticidad: $k_1 = k_4 = 15$ N/m, $k_2 = k_3 = 35$ N/m, y $m_1 = m_2 = m_3 = 1.5$ kg.

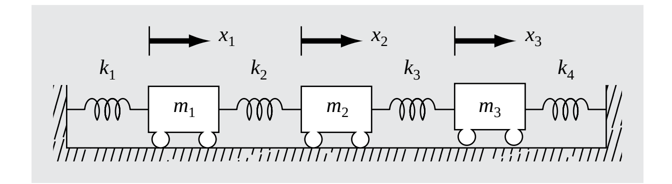


Figura P27.25

27.26 Considere el sistema masa-resorte que se ilustra en la figura P27.26. Las frecuencias para las vibraciones de la masa se determinan con la solución para los valores propios y con la aplicación de $M\ddot{x} + kx = 0$, que da como resultado:

$$\begin{bmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{bmatrix}\begin{Bmatrix} \ddot{x}_1 \\ \ddot{x}_2 \\ \ddot{x}_3 \end{Bmatrix} + \begin{bmatrix} 2k & -k & -k \\ -k & 2k & -k \\ -k & -k & 2k \end{bmatrix}\begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \end{Bmatrix}$$

Al elegir $x = x_0e^{i\omega t}$ como solución, se obtiene la matriz siguiente:

$$\begin{bmatrix} 2k - m_1\omega^2 & -k & -k \\ -k & 2k - m_2\omega^2 & -k \\ -k & -k & 2k - m_3\omega^2 \end{bmatrix}\begin{Bmatrix} x_{01} \\ x_{02} \\ x_{03} \end{Bmatrix} e^{i\omega t} = \begin{Bmatrix} 0 \\ 0 \\ 0 \end{Bmatrix}$$

Utilice el comando `eig` de MATLAB para resolver los valores propios de la matriz anterior $k - m\omega^2$. Después utilice dichos valores

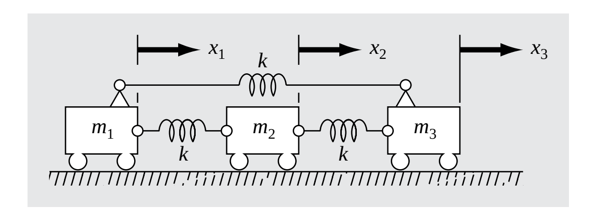


Figura P27.26

propios para resolver para las frecuencias (ω). Haga $m_1 = m_2 = m_3 =$ 1 kg, y k = 2 N/m.

27.27 Hornbeck (1975) propuso la siguiente EDO parásita no lineal:

$$\frac{dy_1}{dt} = 5(y_1 - t^2)$$

Si la condición inicial es $y_1(0) = 0.08$, obtenga una solución de $t = 0$ a 5:

a) Analítica.
b) Con el método de RK de cuarto orden con tamaño de paso constante de 0.03125.
c) Con la función ODE45 de MATLAB.
d) Con la función ODE23S de MATLAB.
e) Con la función ODE23TB de MATLAB.

Presente sus resultados en forma gráfica.

27.28 Una barra caliente con una fuente de calor uniforme se puede modelar con la *ecuación de Poisson*:

$$\frac{d^2T}{dx^2} = -f(x)$$

Dada una fuente de calor $f(x) = 25$ y las condiciones en la frontera $T(x = 0) = 40$ y $T(x = 10) = 200$, resuelva para la distribución de temperatura con *a*) el método del disparo y *b*) el método de diferencias finitas ($\Delta x = 2$).

27.29 Repita el problema 27.28, pero para la siguiente fuente de calor: $f(x) = 0.12x^3 - 2.4x^2 + 12x$.

27.30 Suponga que la posición de un objeto que cae está gobernada por la siguiente ecuación diferencial:

$$\frac{d^2x}{dt^2} + \frac{c}{m}\frac{dx}{dt} - g = 0$$

donde c = un coeficiente de arrastre de primer orden = 12.5 kg/s, m = masa = 70 kg y g = aceleración de la gravedad = 9.81 m/s^2. Use el método de disparo para despejar posición y velocidad de esta ecuación dadas las condiciones de frontera $x(0) = 0$ y $x(12) = 500$.

27.31 Repita el ejemplo 27.3, pero aísle el lado izquierdo de la barra. Es decir, cambie las condiciones de frontera en el lado izquierdo de la barra a $T(0) = 0$.

CAPÍTULO 28

Estudio de casos: ecuaciones diferenciales ordinarias

El propósito de este capítulo es resolver algunas ecuaciones diferenciales ordinarias usando los métodos numéricos presentados en la parte siete. Las ecuaciones provienen de problemas prácticos de la ingeniería. Muchas de estas aplicaciones generan ecuaciones diferenciales no lineales que no se pueden resolver con técnicas analíticas. Por lo tanto, usualmente se requieren métodos numéricos. Así, las técnicas para la solución numérica de ecuaciones diferenciales ordinarias son fundamentales en la práctica de la ingeniería. Los problemas de este capítulo ilustran algunas de las ventajas y desventajas de varios de los métodos desarrollados en la parte siete.

La sección 28.1 plantea un problema en el contexto de la ingeniería química. Ahí se muestra cómo puede simularse el comportamiento transitorio de los reactores químicos. También se ilustra cómo utilizar la optimización para estimar los parámetros de las EDO.

Las secciones 28.2 y 28.3 son tomadas de las ingenierías civil y eléctrica, respectivamente, y tratan con la solución de sistemas de ecuaciones diferenciales. En ambos casos, se necesita gran exactitud y, en consecuencia, se usa el método RK de cuarto orden. Además, el problema de ingeniería eléctrica implica también la determinación de valores propios.

La sección 28.4 emplea varios métodos para investigar el comportamiento de un péndulo oscilante. Este problema también utiliza un sistema de dos ecuaciones diferenciales simultáneas. Un aspecto importante de este ejemplo es que ilustra cómo los métodos numéricos permiten incorporar los efectos no lineales de manera fácil en un análisis de ingeniería.

28.1 USO DE LAS EDO PARA ANALIZAR LA RESPUESTA TRANSITORIA DE UN REACTOR (INGENIERÍA QUÍMICA/BIOINGENIERÍA)

Antecedentes En la sección 12.1 analizamos el estado estacionario de una serie de reactores. Además de los cálculos en estado estacionario, también podríamos estar interesados en la respuesta transitoria de un reactor completamente mezclado. Para ello, desarrollamos expresiones matemáticas para el término de acumulación de la ecuación (12.1).

La acumulación representa el cambio de masa en el reactor por un cambio en el tiempo. En un sistema de volumen constante, esto se formula simplemente como

$$\text{Acumulación} = V\frac{dc}{dt} \tag{28.1}$$

donde V = volumen y c = concentración. Así, una formulación matemática para la acumulación es el volumen por la derivada de c con respecto a t.

En este problema incorporaremos el término acumulación en el balance de masa general que se desarrolló en la sección 12.1. Luego lo utilizaremos para simular la dinámica de un solo reactor y de un sistema de reactores. En el último caso, mostraremos cómo se pueden determinar los valores propios del sistema y analizaremos su dinámica. Por último, se ilustrará cómo se emplea la optimización para estimar los parámetros de los modelos de balance de masa.

Solución Las ecuaciones (28.1) y (12.1) se usan para representar el balance de masa de un solo reactor, como el que se muestra en la figura 28.1:

$$V\frac{dc}{dt} = Qc_{\text{ent}} - Qc \tag{28.2}$$

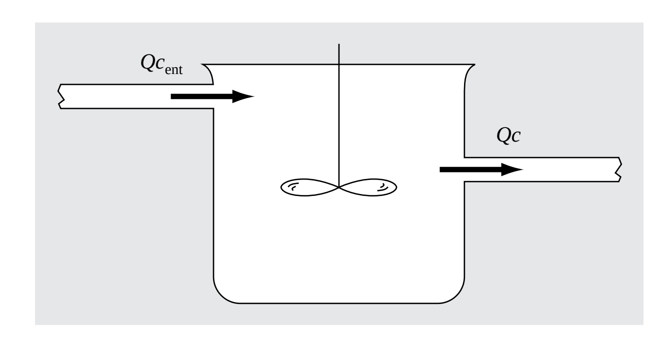


Figura 28.1
Reactor completamente mezclado con un flujo de entrada y un flujo de salida.

Acumulación = entradas – salidas

La ecuación (28.2) se emplea para determinar soluciones transitorias, o variables en el tiempo, para el reactor. Por ejemplo, si $c = c_0$ en $t = 0$, se utiliza el cálculo para obtener en forma analítica la solución de la ecuación (28.2)

$$c = c_{en}(1 - e^{-(Q/V)t}) + c_0 e^{-(Q/V)t}$$

Si $c_{en} = 50$ mg/m^3, $Q = 5$ m^3/min, $V = 100$ m^3 y $c_0 = 10$ mg/m^3, la solución es

$$c = 50(1 - e^{-0.05t}) + 10e^{-0.05t}$$

La figura 28.2 muestra esta solución analítica exacta.

El método de Euler ofrece un procedimiento alternativo para resolver la ecuación (28.2). En la figura 28.2 se presentan dos soluciones con diferentes tamaños de paso. Conforme el tamaño de paso disminuye, la solución numérica converge a la solución analítica. Así, en este caso, el método numérico se utiliza para verificar el resultado analítico.

Además de verificar los resultados dados en forma analítica, las técnicas numéricas son útiles en aquellas situaciones donde las soluciones analíticas son imposibles, o tan difíciles que resultan imprácticas. Por ejemplo, más allá de un solo reactor, los métodos numéricos tienen utilidad para simular la dinámica de sistemas de reactores. Por ejemplo, se pueden escribir EDO para los cinco reactores acoplados de la figura 12.3. El balance de masa para el primer reactor se puede escribir como

$$V_1 \frac{dc_1}{dt} = Q_{01}c_{01} + Q_{31}c_3 - Q_{12}c_1 - Q_{15}c_1$$

o, sustituyendo parámetros (observe que $Q_{01}c_{01} = 50$ mg/min, $Q_{03}c_{03} = 160$ mg/min, $V_1 = 50$ m^3, $V_2 =$ 20 m^3, $V_3 = 40$ m^3, $V_4 = 80$ m^3 y $V_5 = 100$ m^3),

$$\frac{dc_1}{dt} = -0.12c_1 + 0.02c_3 + 1$$

De manera similar, se desarrollan balances para los otros reactores como sigue

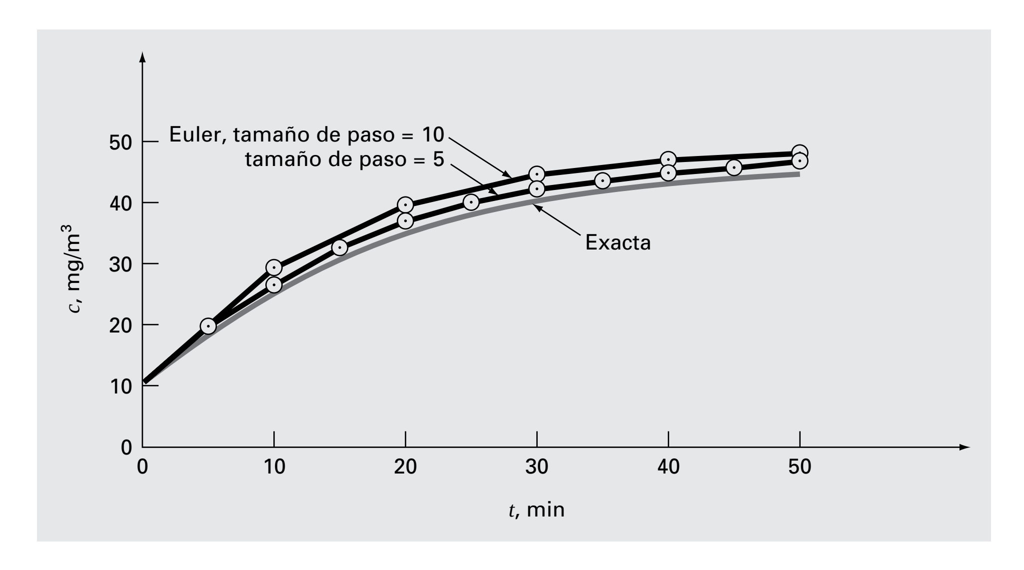


Figura 28.2
Gráfica de las soluciones analítica y numérica de la ecuación (28.2). Las soluciones numéricas se obtienen con el método de Euler usando diferentes tamaños de paso.

$$\frac{dc_2}{dt} = 0.15c_1 - 0.15c_2$$

$$\frac{dc_3}{dt} = 0.025c_2 - 0.225c_3 + 4$$

$$\frac{dc_4}{dt} = 0.1c_3 - 0.1375c_4 + 0.025c_5$$

$$\frac{dc_5}{dt} = 0.03c_1 + 0.01c_2 - 0.04c_5$$

Suponga que en $t = 0$ todas las concentraciones en los reactores son cero. Calcule cómo aumentarán sus concentraciones en la siguiente hora.

Las ecuaciones se integran con el método RK de cuarto orden para un sistema de ecuaciones, y los resultados se ilustran en la figura 28.3. Observe que cada uno de los reactores muestra una respuesta transitoria diferente a la entrada de la sustancia química. Esas respuestas se parametrizan mediante un tiempo de respuesta para 90%, t_{90}, el cual mide el tiempo requerido por cada reactor para alcanzar 90% de su último nivel en estado estacionario. El intervalo de tiempos va desde cerca de 10 minutos en el reactor 3 hasta aproximadamente 70 minutos en el reactor 5. Los tiempos de respuesta de los reactores 4 y 5 son de particular interés, ya que los dos flujos de salida del sistema salen de esos tanques. Así, un ingeniero químico que esté diseñando el sistema podrá cambiar los flujos o volúmenes de los reactores, para acelerar la respuesta de estos tanques manteniendo las salidas deseadas. Los métodos numéricos del tipo que se describen en esta parte del libro son útiles para realizar estos cálculos de diseño.

Una mejor comprensión de las características de respuesta del sistema se obtiene calculando sus valores propios. Primero, el sistema de EDO se escribe como un problema de valores propios:

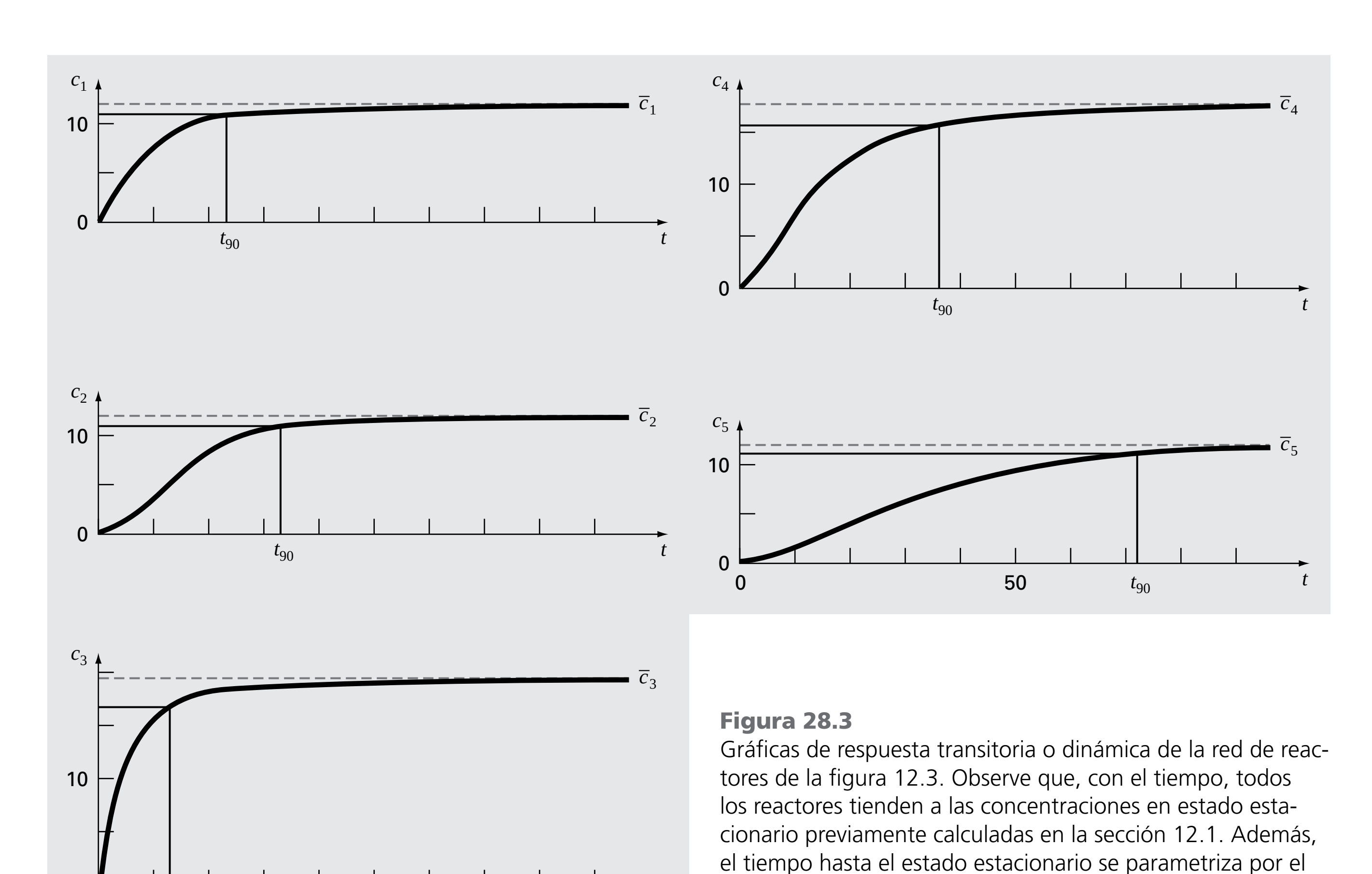


Figura 28.3
Gráficas de respuesta transitoria o dinámica de la red de reactores de la figura 12.3. Observe que, con el tiempo, todos los reactores tienden a las concentraciones en estado estacionario previamente calculadas en la sección 12.1. Además, el tiempo hasta el estado estacionario se parametriza por el tiempo de respuesta para 90%, t_{90}.

$$\begin{bmatrix} 0.12-\lambda & 0 & -0.02 & 0 & 0 \\ -0.15 & 0.15-\lambda & 0 & 0 & 0 \\ 0 & -0.025 & 0.225-\lambda & 0 & 0 \\ 0 & 0 & -0.1 & 0.1375-\lambda & -0.025 \\ -0.03 & -0.01 & 0 & 0 & 0.04-\lambda \end{bmatrix} \begin{Bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{Bmatrix} = \{0\}$$

donde λ y $\{e\}$ = los valores propios y los vectores propios, respectivamente.

Un paquete como MATLAB se utiliza para generar fácilmente los valores propios y los vectores propios.

```
>> a=[0.12 0.0 -0.02 0.0 0.0;-.15 0.15 0.0 0.0 0.0;0.0
-0.025 0.225 0.0 0.0; 0.0 0.0 -.1 0.1375 -0.025; -0.03 -0.01
0.0 0.0 0.04];
>> [e,l]=eig(a)
e =

          0        0   -0.1228  -0.1059   0.2490
          0        0    0.2983   0.5784   0.8444
          0        0    0.5637   0.3041   0.1771
     1.0000   0.2484   -0.7604  -0.7493   0.3675
          0   0.9687    0.0041  -0.0190  -0.2419

l =
     0.1375        0         0        0        0
          0   0.0400         0        0        0
          0        0    0.2118        0        0
          0        0         0   0.1775        0
          0        0         0        0   0.1058
```

Los valores propios se pueden interpretar reconociendo que la solución general de un sistema de EDO se representa como la suma de exponenciales. Por ejemplo, para el reactor 1, la solución general será de la forma

$$c_1 = c_{11}e^{-\lambda_1 t} + c_{12}e^{-\lambda_2 t} + c_{13}e^{-\lambda_3 t} + c_{14}e^{-\lambda_4 t} + c_{15}e^{-\lambda_5 t}$$

donde c_{ij} = la parte de la condición inicial del reactor i que corresponde al j-ésimo valor propio. Así, debido a que, en este caso, todos los valores propios son positivos (y, por lo tanto, negativos en la función exponencial), la solución consiste en un conjunto de exponenciales en decaimiento. Aquel con el valor propio menor (en nuestro caso, 0.04) será el más lento. En algunos casos, el ingeniero que realiza este análisis podrá ser capaz de relacionar este valor propio con los parámetros del sistema. Por ejemplo, el cociente del flujo de salida del reactor 5 entre su volumen es $(Q_{55} + Q_{54})/V_5 = 4/100 = 0.04$. Tal información se utiliza, entonces, para modificar el desempeño de la dinámica del sistema.

Por último quisiéramos revisar en el presente contexto la *estimación del parámetro*. Donde a menudo se presenta lo anterior es en la *cinética de reacción*, es decir, la cuantificación de las velocidades de las reacciones químicas.

Un sencillo ejemplo de esto se ilustra en la figura 28.4. Se tiene un conjunto de matraces que contienen un compuesto químico que decae con el tiempo. A ciertos intervalos de tiempo, se mide y se registra la concentración de uno de los matraces. Así, el resultado es una tabla de tiempos y concentraciones.

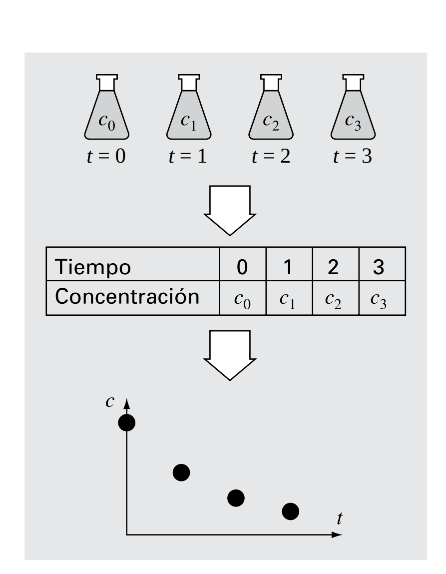


Figura 28.4
Un sencillo experimento para obtener datos de velocidad de un compuesto químico que decae con el tiempo (tomado de Chapra, 1997).

Un modelo comúnmente usado para describir tales datos es

$$\frac{dc}{dt} = -kc^n \tag{28.3}$$

donde k = una velocidad de la reacción y n = el orden de la reacción. Los ingenieros químicos utilizan datos de concentración contra tiempo como los que se presentan en la figura 28.4 para estimar k y n. Una manera de hacerlo es suponer valores de los parámetros y después resolver numéricamente la ecuación (28.3). Los valores pronosticados de concentración se comparan con las concentraciones medidas y se realiza la valoración del ajuste. Si el ajuste se considera inadecuado (por ejemplo, al examinar una gráfica o una medición estadística como la suma de los cuadrados de los residuos), los valores pronosticados se ajustan y se repite el procedimiento hasta que se alcanza un ajuste apropiado.

Los siguientes datos se ajustan de la siguiente manera:

t, d	0	1	3	5	10	15	20
c, mg/L	12	10.7	9	7.1	4.6	2.5	1.8

La solución a este problema se muestra en la figura 28.5. Se utilizó una hoja de cálculo de Excel para realizar el cálculo.

	A	B	C	D	E	F	G	H
1	Ajuste de la velocidad de reacción							
2	Datos con la integral/procedimiento de mínimos cuadrados							
3	k	0.091528						
4	n	1.044425						
5	dt	1						
6	t	k1	k2	k3	k4	cp	cm	(cp-cm)^2
7	0	−1.22653	−1.16114	−1.16462	−1.10248	12	12	0
8	1	−1.10261	−1.04409	−1.04719	−0.99157	10.83658	10.7	0.018653
9	2	−0.99169	−0.93929	−0.94206	−0.89225	9.790448		
10	3	−0.89235	−0.84541	−0.84788	−0.80325	8.849344	9	0.022697
11	4	−0.80334	−0.76127	−0.76347	−0.72346	8.002317		
12	5	−0.72354	−0.68582	−0.68779	−0.65191	7.239604	7.1	0.019489
13	6	−0.65198	−0.61814	−0.61989	−0.5877	6.552494		
14	7	−0.58776	−0.55739	−0.55895	−0.53005	5.933207		
15	8	−0.53011	−0.50283	−0.50424	−0.47828	5.374791		
16	9	−0.47833	−0.45383	−0.45508	−0.43175	4.871037		
17	10	−0.4318	−0.40978	−0.4109	−0.38993	4.416389	4.6	0.033713
18	11	−0.38997	−0.37016	−0.37117	−0.35231	4.005877		
19	12	−0.35234	−0.33453	−0.33543	−0.31846	3.635053		
20	13	−0.31849	−0.30246	−0.30326	−0.28798	3.299934		
21	14	−0.28801	−0.27357	−0.2743	−0.26054	2.996949		
22	15	−0.26056	−0.24756	−0.24821	−0.23581	2.7229	2.5	0.049684
23	16	−0.23583	−0.22411	−0.22469	−0.21352	2.474917		
24	17	−0.21354	−0.20297	−0.20349	−0.19341	2.250426		
25	18	−0.19343	−0.18389	−0.18436	−0.17527	2.047117		
26	19	−0.17529	−0.16668	−0.16711	−0.1589	1.862914		
27	20	−0.15891	−0.15115	−0.15153	−0.14412	1.695953	1.8	0.010826
28								
29							SSR =	0.155062

Figura 28.5
Aplicación de una hoja de cálculo y de métodos numéricos para determinar el orden y el coeficiente de la velocidad de reacción con los datos. Esta tabla se obtuvo con la hoja de cálculo Excel.

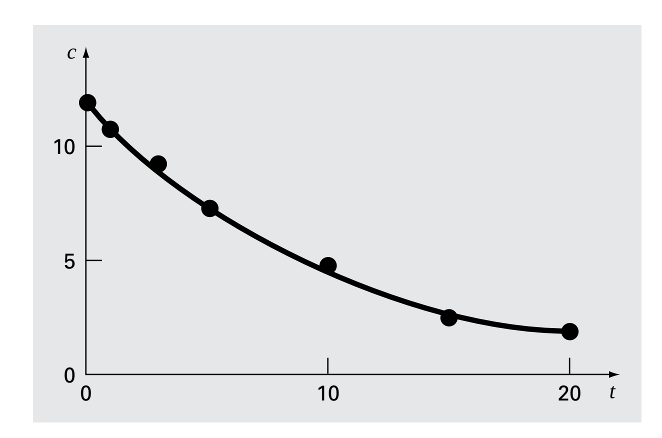


Figura 28.6
Gráfica del ajuste generado con el método de la integral/de mínimos cuadrados.

Se introducen valores iniciales de la velocidad de reacción y del orden de reacción en las celdas B3 y B4, respectivamente, el tamaño de paso para el cálculo numérico se teclea en la celda B5. En este caso, se introduce una columna con el tiempo de los cálculos en la columna A, comenzando en 0 (celda A7) y terminando en 20 (celda A27). Los coeficientes desde k_1 hasta k_4 del método RK de cuarto orden se calculan en el bloque B7..E27. Éstos se usan ahora para determinar las predicciones de las concentraciones (los valores c_p) en la columna F. Los valores medidos (c_m) se introducen en la columna G adyacente a la columna de predicciones correspondientes, las cuales se utilizan después, conjuntamente, para calcular las diferencias elevadas al cuadrado en la columna H. Por último, estos valores se suman en la celda H29.

Aquí, el Solver de Excel se utiliza para determinar los mejores valores de los parámetros. Una vez que usted haya entrado al Solver, se le pide una celda objetivo o solución (H29); además se le pregunta si usted quiere maximizar o minimizar la celda objetivo (minimizar), y que dé las celdas que se van a variar (B3..B4). Después, usted activa el algoritmo [**s**(olve)], y los resultados son como los de la figura 28.5. Como se muestra, los valores de las celdas B3..B4 ($k = 0.0915$ y $n = 1.044$) minimizan la suma de los cuadrados de los residuales (SSR = 0.155) entre los datos de predicción y los datos medidos. En la figura 28.6 se presenta una gráfica del ajuste junto con los datos.

28.2 MODELOS DEPREDADOR-PRESA Y CAOS (INGENIERÍA CIVIL/AMBIENTAL)

Antecedentes Los ingenieros ambientales modelan diversos problemas que implican sistemas de ecuaciones diferenciales ordinarias no lineales. En esta sección nos concentraremos en dos de estos problemas. El primero se relaciona con los modelos llamados depredador-presa, que se utilizan en el estudio de ciclos de nutrientes y contaminantes tóxicos en las cadenas alimenticias acuáticas, y de sistemas de tratamiento biológicos. El segundo son ecuaciones obtenidas de la dinámica de fluidos, que se utilizan para simular la atmósfera. Además de sus obvias aplicaciones en el pronóstico del tiempo, tales ecuaciones también son útiles para estudiar la contaminación del aire y el cambio climático mundial.

Los *modelos depredador-presa* se desarrollaron de manera independiente en la primera parte del siglo XX, gracias al trabajo del matemático italiano Vito Volterra y del biólogo estadounidense Alfred J. Lotka. Estas ecuaciones se conocen como las *ecuaciones de Lotka-Volterra*. El ejemplo más simple es el siguiente sistema de EDO:

$$\frac{dx}{dt} = ax - bxy \tag{28.4}$$

$$\frac{dy}{dt} = -cy + dxy \tag{28.5}$$

donde x y y = número de presas y depredadores, respectivamente, a = la razón de crecimiento de la presa, c = la razón de muerte del depredador, y b y d = la razón que caracteriza el efecto de la interacción depredador-presa sobre la muerte de la presa y el crecimiento del depredador, respectivamente. Los términos que se multiplican (es decir, los que involucran xy) hacen que las ecuaciones sean no lineales.

Un ejemplo de un modelo sencillo basado en las dinámicas del fluido atmosférico son las *ecuaciones de Lorenz*, desarrolladas por el meteorólogo estadounidense Edward Lorenz,

$$\frac{dx}{dt} = -\sigma x + \sigma y \tag{28.6}$$

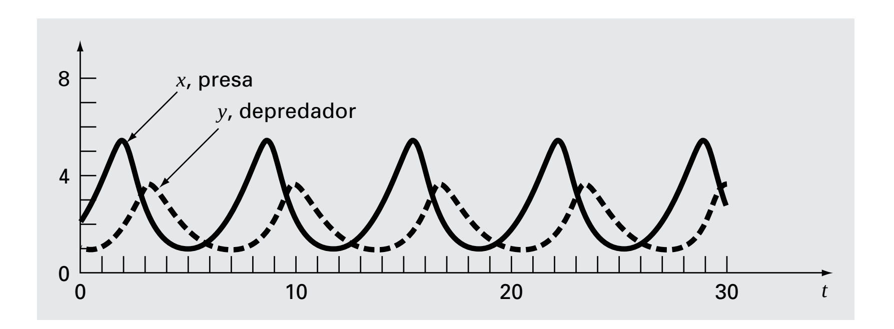


Figura 28.7
Representación en el dominio del tiempo de los números de presas y depredadores con el modelo de Lotka-Volterra.

$$\frac{dy}{dt} = rx - y - xz \tag{28.7}$$

$$\frac{dz}{dt} = -bz + xy \tag{28.8}$$

Lorenz desarrolló esas ecuaciones para relacionar la intensidad del movimiento de fluido atmosférico, x, con las variaciones de temperatura y y z en las direcciones horizontal y vertical, respectivamente. Como en el modelo depredador-presa, observamos que la no linealidad está dada por los términos que se multiplican (xz y xy).

Use métodos numéricos para obtener las soluciones de estas ecuaciones. Grafique los resultados para visualizar cómo las variables dependientes cambian en el tiempo. Además, grafique las variables dependientes una contra otra para observar si surge algún patrón interesante.

Solución Utilice los siguientes valores de los parámetros para la simulación depredador-presa: a = 1.2, b = 0.6, c = 0.8 y d = 0.3. Emplee como condiciones iniciales x = 2 y y = 1 e integre desde t = 0 hasta 30. Usaremos el método RK de cuarto orden con doble precisión para obtener las soluciones.

En la figura 28.7 se muestran los resultados usando un tamaño de paso de 0.1. Advierta que surge un patrón cíclico. Así, como inicialmente la población del depredador es pequeña, la presa crece de manera exponencial. En cierto momento, las presas son tan numerosas que la población del depredador empieza a crecer. Después el aumento de depredadores causa que la presa disminuya. Esta disminución, a su vez, lleva a una disminución de los depredadores. Con el tiempo, el proceso se repite. Observe que, como se esperaba, el pico en la curva para el depredador se retrasa respecto al de la presa. Además, observe que el proceso tiene un periodo fijo; es decir, se repite cada cierto tiempo.

Ahora, si se cambiaran los parámetros usados para simular la figura 28.7, aunque el patrón general seguirá siendo el mismo, las magnitudes de los picos, retrasos y periodos cambiarán. Así, existe un número infinito de ciclos que podrán ocurrir.

Una representación del espacio-fase es útil para distinguir la estructura fundamental del modelo. En lugar de graficar x y y contra t, se grafica x contra y. Esta gráfica ilustra la manera en que interactúan las variables de estado (x y y) y se le conoce como una *representación del espacio-fase.*

En la figura 28.8 se muestra la representación estado-espacio del caso que estamos estudiando. La interacción entre el depredador y la presa define una órbita cerrada en sentido derecho.

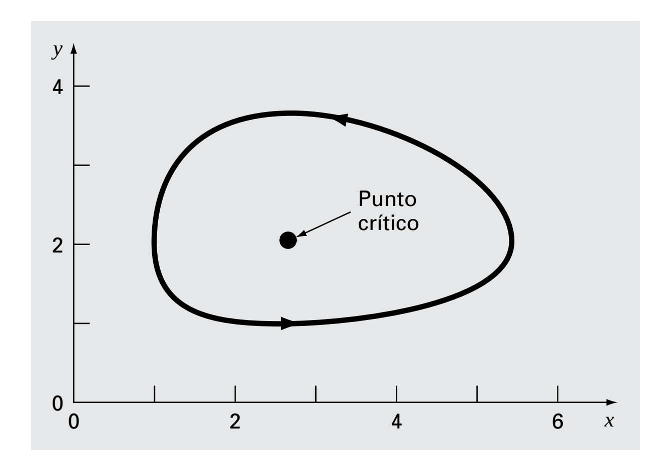


Figura 28.8
Representación del espacio-fase del modelo Lotka-Volterra.

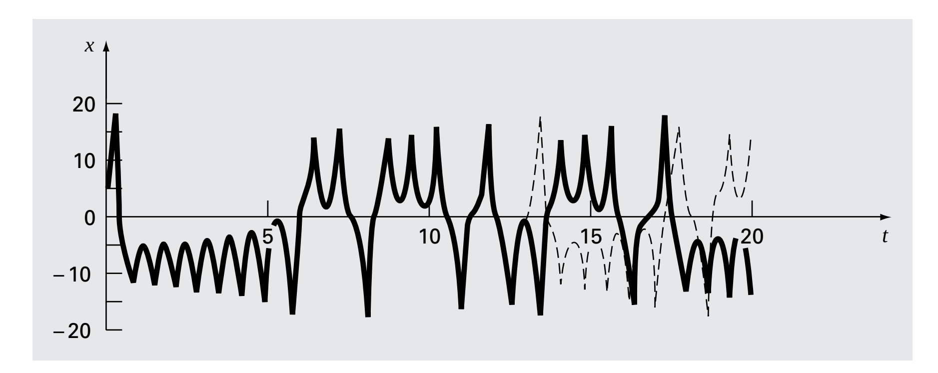


Figura 28.9
Representación en el dominio del tiempo de x contra t para las ecuaciones de Lorenz. La línea sólida para la serie del tiempo es para las condiciones iniciales (5, 5, 5). La línea punteada es cuando la condición inicial para x está ligeramente perturbada (5.001, 5, 5).

Observe que hay un punto crítico o de reposo en el centro de la órbita. La localización exacta de este punto se determina poniendo las ecuaciones (28.4) y (28.5) en estado estacionario ($dy/dt = dx/dt = 0$) y resolviendo para $(x, y) = (0, 0)$ y $(c/d, a/b)$. La primera es el resultado trivial, si empezamos sin depredador y sin presa, no sucederá nada. La segunda es el resultado más interesante si las condiciones iniciales se consideran como $x = c/d$ y $y = a/b$, la derivada será cero y las poblaciones permanecerán constantes.

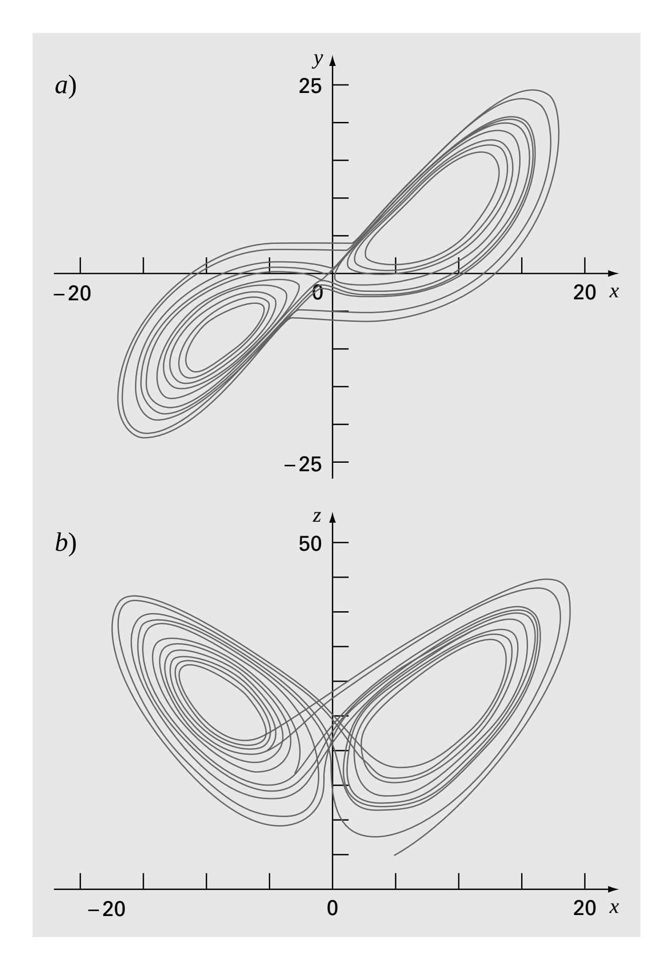


Figura 28.10
Representación del estado-fase de las ecuaciones de Lorenz. *a*) Proyección *xy*; *b*) proyección *xz*.

Ahora, utilicemos el mismo procedimiento para investigar el comportamiento de las ecuaciones de Lorenz con los siguientes valores de los parámetros: $\sigma = 10$, $b = 2.666667$ y $r = 28$. Emplee como condiciones iniciales $x = y = z = 5$, e integre desde $t = 0$ hasta 20. De nuevo, usaremos el método RK de cuarto orden con doble precisión para obtener las soluciones.

Los resultados mostrados en la figura 28.9 son muy diferentes al comportamiento de las ecuaciones de Lotka-Volterra. La variable x parece experimentar un patrón casi aleatorio de oscilaciones, rebotando de valores negativos a positivos. Sin embargo, aun cuando los patrones parezcan aleatorios, la frecuencia de la oscilación y las amplitudes parecen bastante consistentes.

Otra característica interesante se puede ilustrar cambiando ligeramente la condición inicial de x (de 5 a 5.001). Los resultados están superpuestos como una línea punteada en la figura 28.9. Aunque las soluciones siguen un mismo comportamiento por un tiempo, después de aproximadamente $t = 12.5$ divergen significativamente. Así, se observa que las ecuaciones de Lorenz son muy sensibles a las condiciones iniciales. En su estudio original, esto llevó a Lorenz a la conclusión de que ¡pronosticar el clima a largo plazo será imposible!

Por último, examinemos las gráficas del estado-fase. Como tenemos tres variables independientes, estamos limitados a proyecciones. En la figura 28.10 se muestran las proyecciones en los planos *xy* y *xz*. Observe que se manifiesta una estructura cuando la percibimos desde la perspectiva estado-fase. La solución forma órbitas alrededor de lo que parecen ser puntos críticos. Dichos puntos se llaman *atractores extraños*, en la jerga de los matemáticos que estudian tales sistemas no lineales.

A las soluciones del tipo que hemos explorado en las ecuaciones de Lorenz se les conoce como soluciones *caóticas*. Actual-

mente, el estudio del caos y de los sistemas no lineales representa una interesante área del análisis que tiene implicaciones tanto en las matemáticas como en la ciencia y en la ingeniería.

Desde una perspectiva numérica, el punto principal es la sensibilidad de tales soluciones a las condiciones iniciales. Así, los diferentes algoritmos numéricos, la precisión de la computadora y la determinación del tamaño de paso tienen un impacto sobre la solución numérica que se obtenga.

28.3 SIMULACIÓN DE LA CORRIENTE TRANSITORIA EN UN CIRCUITO ELÉCTRICO (INGENIERÍA ELÉCTRICA)

Antecedentes Son comunes los circuitos eléctricos en los que la corriente varía con el tiempo, en lugar de permanecer constante. Cuando se cierra súbitamente el interruptor, se establece una corriente transitoria en el lado derecho del circuito que se muestra en la figura 28.11.

Las ecuaciones que describen el comportamiento transitorio del circuito de la figura 28.11 se basan en las leyes de Kirchhoff, que establecen que la suma algebraica de las caídas de voltaje alrededor de un ciclo cerrado es cero (recuerde la sección 8.3). Así,

$$L\frac{di}{dt} + Ri + \frac{q}{c} - E(t) = 0 \tag{28.9}$$

donde $L(di/dt)$ = la caída de voltaje a través del inductor, L = inductancia (H), R = resistencia (Ω), q = carga en el capacitor (C), C = capacitancia (F), $E(t)$ = fuente de voltaje variable en el tiempo (V), e

$$i = \frac{dq}{dt} \tag{28.10}$$

Las ecuaciones (28.9) y (28.10) son un sistema de ecuaciones diferenciales lineales de primer orden que se pueden resolver analíticamente. Por ejemplo, si $E(t) = E_0$ sen ωt y $R = 0$,

$$q(t) = \frac{-E_0}{L(p^2 - \omega^2)}\frac{\omega}{p}\text{sen } pt + \frac{E_0}{L(p^2 - \omega^2)}\text{sen }\omega t \tag{28.11}$$

donde $p = 1/\sqrt{LC}$. Los valores de q y dq/dt son cero para $t = 0$. Utilice un procedimiento numérico para resolver las ecuaciones (28.9) y (28.10), y compare los resultados con la ecuación (28.11).

Solución Este problema comprende un intervalo de integración bastante amplio y requiere de un esquema de gran exactitud para resolver la ecuación diferencial, si se esperan buenos resultados. Supongamos que $L = 1$ H, $E_0 = 1$ V, $C = 0.25$ C y $\omega^2 = 3.5$ s^2. Esto da $p = 2$, y la ecuación (28.11) se convierte en

$$q(t) = -1.8708 \text{ sen } (2t) + 2 \text{ sen } (1.8708t)$$

para la solución analítica. La gráfica de esta función se muestra en la figura 28.12. La naturaleza cambiante de la función exige necesariamente de un procedimiento numérico para encontrar

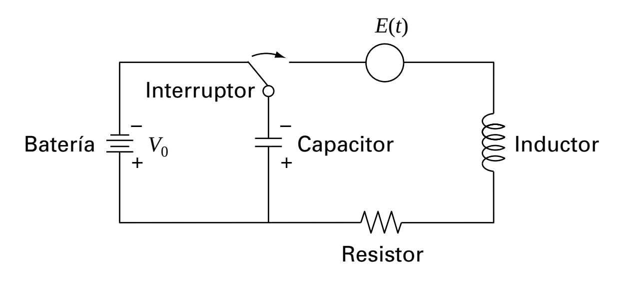


Figura 28.11
Circuito eléctrico donde la corriente varía con el tiempo.

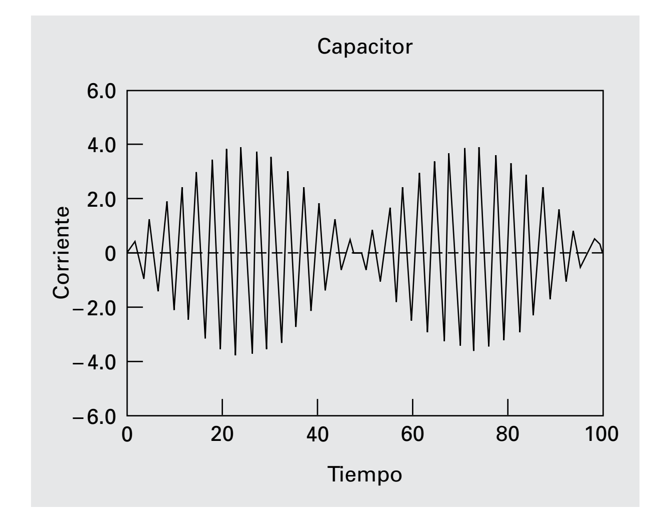


Figura 28.12
Pantalla de computadora donde se muestra la gráfica de la función obtenida en la ecuación (28.11).

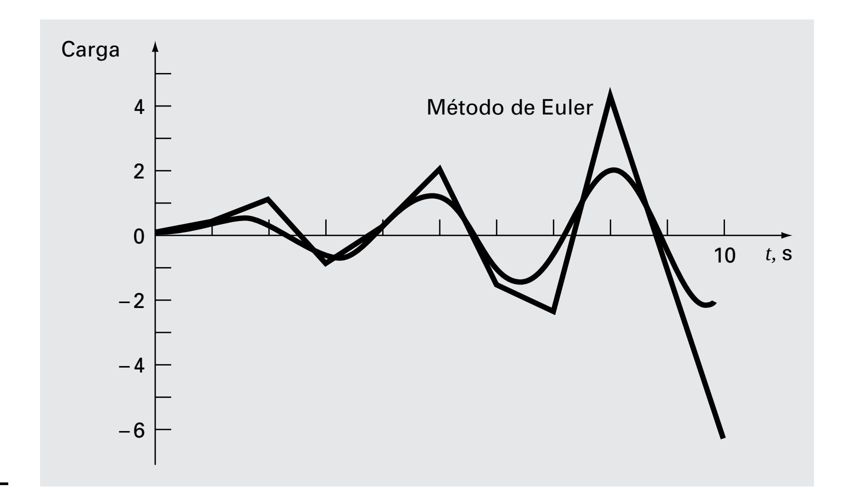


Figura 28.13
Resultados de la integración de Euler contra la solución exacta. Observe que sólo se grafica cada décimo punto de salida.

$q(t)$. Además, como la función exhibe una naturaleza periódica que varía lentamente, así como una variación rápida, se necesitan intervalos de integración largos para encontrar la solución. Por estas razones se espera que un método de orden superior sea el adecuado para este problema.

Sin embargo, podemos probar tanto el método de Euler como el RK de cuarto orden y comparar los resultados. Usando un tamaño de paso de 0.1 s, se obtiene un valor de q en t = 10 s de –6.638, con el método de Euler; y un valor de –1.9897, con el método RK de cuarto orden. Estos resultados se comparan con una solución exacta de –1.996 C.

En la figura 28.13 se muestran los resultados de la integración de Euler cada 1.0 s comparada con la solución exacta. Observe que sólo se grafica cada décimo punto de salida. Note que el error global aumenta conforme t aumenta. Este comportamiento divergente se intensifica conforme t se aproxima al infinito.

Además, para simular directamente una respuesta transitoria de una red, los métodos numéricos también se utilizan para determinar sus valores propios. Por ejemplo, en la figura 28.14 se muestra un circuito LC para el cual puede emplearse la ley de voltaje de Kirchhoff para desarrollar el siguiente sistema de EDO:

$$-L_1 \frac{di_1}{dt} - \frac{1}{C_1}\int_{-\infty}^{t} (i_1 - i_2)\, dt = 0$$

$$-L_2 \frac{di_2}{dt} - \frac{1}{C_2}\int_{-\infty}^{t} (i_2 - i_3)\, dt + \frac{1}{C_1}\int_{-\infty}^{t} (i_1 - i_2)\, dt = 0$$

$$-L_3 \frac{di_3}{dt} - \frac{1}{C_3}\int_{-\infty}^{t} i_3\, dt + \frac{1}{C_2}\int_{-\infty}^{t} (i_2 - i_3)\, dt = 0$$

Observe que hemos representado la caída de voltaje a través del capacitor como

$$V_C = \frac{1}{C}\int_{-\infty}^{t} i\, dt$$

Ésta es una expresión alternativa y equivalente a la relación usada en la ecuación (28.9), que se presentó en la sección 8.3.

El sistema de EDO se deriva y se reordena para llegar a

$$L_1 \frac{d^2 i_1}{dt^2} + \frac{1}{C_1}(i_1 - i_2) = 0$$

$$L_2 \frac{d^2 i_2}{dt^2} + \frac{1}{C_2}(i_2 - i_3) - \frac{1}{C_1}(i_1 - i_2) = 0$$

$$L_3 \frac{d^2 i_3}{dt^2} + \frac{1}{C_3} i_3 - \frac{1}{C_2}(i_2 - i_3) = 0$$

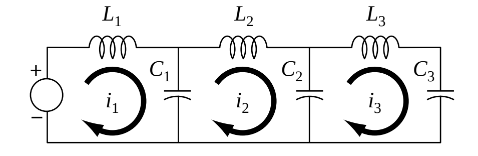


Figura 28.14
Un circuito LC.

La comparación de este sistema con el de la ecuación (27.5) indica una analogía entre un sistema masa-resorte y un circuito *LC*. Como se hizo con la ecuación (27.5), la solución se considera de la forma

$$i_j = A_j \text{ sen } (\omega t)$$

Esta solución, junto con su segunda derivada, se sustituye en las EDO simultáneas. Después de simplificar, el resultado es

$$\begin{aligned}
\left(\frac{1}{C_1} - L_1\omega^2\right)A_1 - \frac{1}{C_2}A_2 &= 0\\
-\frac{1}{C_1}A_1 + \left(\frac{1}{C_1} + \frac{1}{C_2} - L_2\omega^2\right)A_2 - \frac{1}{C_2}A_3 &= 0\\
-\frac{1}{C_2}A_2 + \left(\frac{1}{C_2} + \frac{1}{C_3} - L_3\omega^2\right)A_3 &= 0
\end{aligned}$$

Así, hemos formulado un problema de valores propios. Al hacer una simplificación se tiene el caso especial donde las *C* y las *L* son constantes. En dicha situación, el sistema se expresa en forma matricial como

$$\begin{bmatrix} 1-\lambda & -1 & 0 \\ -1 & 2-\lambda & -1 \\ 0 & -1 & 2-\lambda \end{bmatrix} \begin{Bmatrix} A_1 \\ A_2 \\ A_3 \end{Bmatrix} = \{0\} \tag{28.12}$$

donde

$$\lambda = LC\omega^2 \tag{28.13}$$

Se pueden emplear métodos numéricos para determinar los valores y vectores propios. MATLAB resulta particularmente conveniente para este cálculo. Así, se desarrolló la siguiente sesión en MATLAB para realizar esto:

```
>>a=[1 -1 0; -1 2 -1; 0 -1 2]
a =
     1   -1    0
    -1    2   -1
     0   -1    2
>>[v,d]=eig(a)
v =
    0.7370    0.5910    0.3280
    0.5910   -0.3280   -0.7370
    0.3280   -0.7370    0.5910
d =
    0.1981         0         0
         0    1.5550         0
         0         0    3.2470
```

La matriz v contiene los tres vectores propios del sistema (ordenados en columnas), y *d* es una matriz con los correspondientes valores propios en la diagonal. Así, en el paquete se calcula que los valores propios son: $\lambda = 0.1981$, 1.555 y 3.247. Estos valores, a su vez, pueden sustituirse en la ecuación (28.13) para encontrar las frecuencias naturales del sistema

$$\omega = \begin{cases} \dfrac{0.4450}{\sqrt{LC}} \\ \dfrac{1.2470}{\sqrt{LC}} \\ \dfrac{1.8019}{\sqrt{LC}} \end{cases}$$

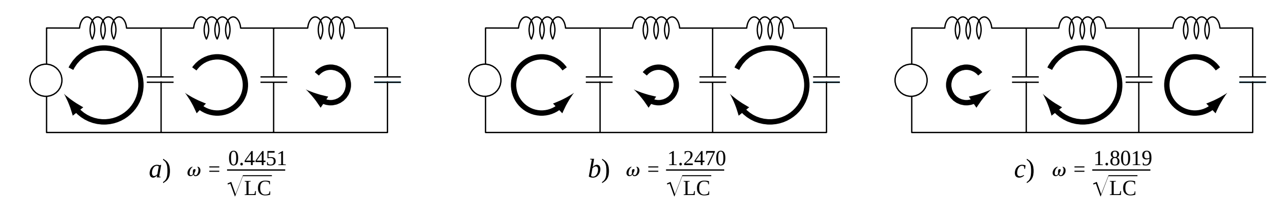


Figura 28.15
Representación visual de los modos de oscilación naturales del circuito *LC* de la figura 28.14.
Observe que los diámetros de las flechas circulares son proporcionales a las magnitudes de las corrientes en cada ciclo.

Además de proporcionar las frecuencias naturales, los valores propios se sustituyen en la ecuación (28.12) para saber más acerca del comportamiento físico del circuito. Por ejemplo, sustituyendo $\lambda = 0.1981$ se obtiene

$$\begin{bmatrix} 0.8019 & -1 & 0 \\ -1 & 1.8019 & -1 \\ 0 & -1 & 1.8019 \end{bmatrix} \begin{Bmatrix} i_1 \\ i_2 \\ i_3 \end{Bmatrix} = \{0\}$$

Como este sistema no tiene una solución única, las corrientes están relacionadas de la siguiente manera

$$0.8019 i_1 = i_2 = 1.8019 i_3 \tag{28.14}$$

Así, como se ilustra en la figura 28.15*a*, oscilan en la misma dirección con diferentes magnitudes. Observe que si suponemos que $i_1 = 0.737$, entonces utilizamos la ecuación (28.14) para calcular el valor de las otras corrientes con el siguiente resultado:

$$\{i\} = \begin{Bmatrix} 0.737 \\ 0.591 \\ 0.328 \end{Bmatrix}$$

que es la primera columna de la matriz v calculada con MATLAB.

De manera similar, al sustituir el segundo valor propio $\lambda = 1.555$, el resultado será

$$-1.8018 i_1 = i_2 = 2.247 i_3$$

Como se ilustra en la figura 28.15*b*, el primer ciclo oscila en dirección opuesta respecto al segundo y al tercero. Por último, el tercer modo se determina como

$$-0.445 i_1 = i_2 = -0.8718 i_3$$

En consecuencia, como se muestra en la figura 28.15*c*, el primero y el tercer ciclos oscilan en dirección opuesta al segundo.

28.4 EL PÉNDULO OSCILANTE (INGENIERÍA MECÁNICA/AERONÁUTICA)

Antecedentes Los ingenieros mecánicos (así como todos los otros ingenieros) a menudo enfrentan problemas relacionados con el movimiento periódico de cuerpos libres. Para abordar tales problemas se requiere conocer la posición y la velocidad de un cuerpo en función del tiempo. Tales funciones son invariablemente la solución de ecuaciones diferenciales ordinarias. Estas ecuaciones diferenciales se basan en las leyes del movimiento de Newton.

Como ejemplo sencillo, considere el péndulo simple que se presentó en la figura PT7.1. La partícula de peso *W* está suspendida de un cable sin peso de longitud *l*. Las únicas fuerzas que actúan

sobre esta partícula son su peso y la tensión R en el cable. La posición de la partícula en cualquier instante está completamente especificada en términos del ángulo θ y l.

El diagrama de cuerpo libre de la figura 28.16 muestra las fuerzas que actúan sobre la partícula y la aceleración. Es conveniente aplicar las leyes del movimiento de Newton en la dirección x, tangente a la trayectoria de la partícula:

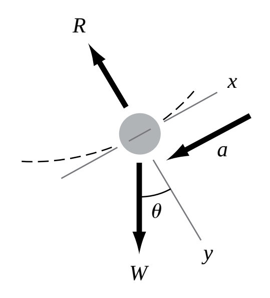


Figura 28.16 Diagrama de cuerpo libre del péndulo oscilante, donde se muestran las fuerzas sobre la partícula y la aceleración.

$$\Sigma F = -W \text{ sen } \theta = \frac{W}{g} a$$

donde g = la constante gravitacional (32.2 ft/s^2) y a = la aceleración en la dirección x. La aceleración angular de la partícula (α) es

$$\alpha = \frac{a}{l}$$

Por lo tanto, en coordenadas polares ($\alpha = d^2\theta/dt^2$),

$$-W \text{ sen } \theta = \frac{Wl}{g}\alpha = \frac{Wl}{g}\frac{d^2\theta}{dt^2}$$

o

$$\frac{d^2\theta}{dt^2} + \frac{g}{l} \text{ sen } \theta = 0 \tag{28.15}$$

Esta ecuación aparentemente simple es una ecuación diferencial no lineal de segundo orden. En general, es difícil o imposible resolver tales ecuaciones de manera analítica. Usted tiene dos opciones para poder seguir adelante. Primero, reducir la ecuación diferencial a una forma que sea posible resolver analíticamente (recuerde la sección PT7.1.1) o, segundo, utilizar una técnica de aproximación numérica para resolver la ecuación diferencial de manera directa. Examinaremos ambas opciones en este ejemplo.

Solución Procediendo con la primera opción, recordemos que la expansión en series de potencias para sen θ está dada por

$$\text{sen } \theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \cdots \tag{28.16}$$

Para desplazamientos angulares pequeños, sen θ es aproximadamente igual a θ cuando se expresa en radianes. Por lo tanto, para desplazamientos pequeños, la ecuación (28.15) se convierte en

$$\frac{d^2\theta}{dt^2} + \frac{g}{l}\theta = 0 \tag{28.17}$$

la cual es una ecuación diferencial lineal de segundo orden. Esta aproximación es muy importante, ya que la ecuación (28.17) es fácil de resolver analíticamente. La solución, basada en la teoría de las ecuaciones diferenciales, está dada por

$$\theta(t) = \theta_0 \cos \sqrt{\frac{g}{l}} t \tag{28.18}$$

donde θ_0 = el desplazamiento en $t = 0$ y donde se supone que la velocidad ($v = d\theta/dt$) es cero en $t = 0$. Al tiempo requerido por el péndulo para un ciclo completo de oscilación se le llama *periodo*, y está dado por

$$T = 2\pi \sqrt{\frac{l}{g}} \tag{28.19}$$

En la figura 28.17 se muestra una gráfica del desplazamiento θ y la velocidad $d\theta/dt$ en función del tiempo, obtenidas a partir de la ecuación (28.18) con $\theta_0 = \pi/4$ y $l = 2$ ft. El periodo, como se calculó con la ecuación (28.19), es 1.5659 s.

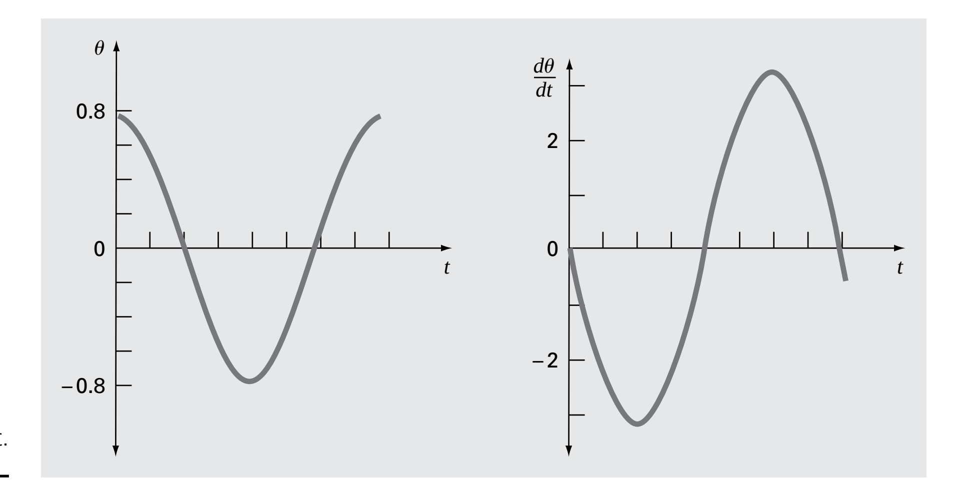


Figura 28.17
Gráfica del desplazamiento θ y la velocidad $d\theta/dt$ en función del tiempo t, como se calculó a partir de la ecuación (28.18). θ_0 es $\pi/4$ y la longitud es de 2 ft.

Los cálculos anteriores son, esencialmente, una solución completa del movimiento del péndulo. Sin embargo, usted debe considerar también la exactitud de los resultados debido a las suposiciones inherentes en la ecuación (28.17). Para evaluar la exactitud, es necesario obtener una solución numérica de la ecuación (28.15), que es una representación física más completa del movimiento. Cualquiera de los métodos analizados en los capítulos 25 y 26 podrán utilizarse con tal propósito (por ejemplo, los métodos de Euler y RK de cuarto orden). La ecuación (28.15) se debe transformar en un sistema de dos ecuaciones de primer orden para que sea compatible con los métodos anteriores. Esto se lleva a cabo de la siguiente manera. La velocidad v está definida por

$$\frac{d\theta}{dt} = v \tag{28.20}$$

y, por lo tanto, la ecuación (28.15) se expresa como

$$\frac{dv}{dt} = -\frac{g}{l}\operatorname{sen}\theta \tag{28.21}$$

Las ecuaciones (28.20) y (28.21) constituyen un sistema de dos ecuaciones diferenciales ordinarias. Las soluciones numéricas utilizando los métodos de Euler y RK de cuarto orden dan los resultados que se muestran en la tabla 28.1, que son semejantes a la solución analítica de la ecuación lineal del movimiento [ecuación (28.18)] de la columna *a*), con las soluciones numéricas de las columnas *b*), *c*) y *d*).

Los métodos de Euler y RK de cuarto orden dan resultados diferentes y ninguno de ellos concuerda con la solución analítica; aunque el método RK de cuarto orden en el caso no lineal es más cercano a la solución analítica que el método de Euler. Para evaluar adecuadamente la diferencia entre los modelos lineal y no lineal, es importante determinar la exactitud de los resultados numéricos. Esto se lleva a cabo de tres maneras. Primero, se reconoce fácilmente que la solución numérica de Euler es inadecuada debido a que sobrepasa la condición inicial en $t = 0.8$ s. Esto viola claramente la conservación de la energía. Segundo, las columnas *c*) y *d*) de la tabla 28.1 muestran la solución del método RK de cuarto orden con tamaños de paso 0.05 y 0.01. Como éstas varían en la cuarta cifra decimal, es razonable suponer que la solución con un tamaño de paso de 0.01 sea también exacta con este grado de certeza. Tercero, en el caso con tamaño de paso de 0.01, θ tiene un valor máximo local de 0.785385 en $t = 1.63$ s (no mostrado en la tabla 28.1). Esto indica que el péndulo regresa a su posición original, con una exactitud de cuatro cifras, en un periodo de 1.63 s. Estas consideraciones le permiten suponer con seguridad que la diferencia entre las columnas *a*) y *d*) de la tabla 28.1 representa verdaderamente la diferencia entre el modelo lineal y el no lineal.

Otra forma de caracterizar la diferencia entre el modelo lineal y el no lineal se basa en el periodo. En la tabla 28.2 se indica el periodo de oscilación, como se calculó con los modelos lineal y no lineal para tres diferentes desplazamientos iniciales. Se aprecia que los periodos calculados concuerdan bastante cuando θ es pequeña, ya que θ es una buena aproximación para sen θ en la ecuación (28.16). Esta aproximación se deteriora cuando θ se vuelve grande.

Tabla 28.1 Comparación de una solución analítica lineal del problema del péndulo oscilante, con tres soluciones numéricas no lineales.

		Soluciones numéricas no lineales		
Tiempo, s	Solución analítica lineal *a*)	Euler (h = 0.05) *b*)	RK de cuarto orden (h = 0.05) *c*)	RK de cuarto orden (h = 0.01) *d*)
0.0	0.785398	0.785398	0.785398	0.785398
0.2	0.545784	0.615453	0.566582	0.566579
0.4	–0.026852	0.050228	0.021895	0.021882
0.6	–0.583104	–0.639652	–0.535802	–0.535820
0.8	–0.783562	–1.050679	–0.784236	–0.784242
1.0	–0.505912	–0.940622	–0.595598	–0.595583
1.2	0.080431	–0.299819	–0.065611	–0.065575
1.4	0.617698	0.621700	0.503352	0.503392
1.6	0.778062	1.316795	0.780762	0.780777

Tabla 28.2 Comparación del periodo de un cuerpo oscilante, calculado con los modelos lineal y no lineal.

	Periodo, s	
Desplazamiento inicial, θ_0	Modelo lineal ($T = 2\pi\sqrt{1/g}$)	Modelo no lineal [solución numérica de la ecuación (28.15)]
$\pi/16$	1.5659	1.57
$\pi/4$	1.5659	1.63
$\pi/2$	1.5659	1.85

Estos análisis son típicos de los casos que usted encontrará cotidianamente como ingeniero. La utilidad de las técnicas numéricas se vuelve particularmente importante en problemas no lineales y, en muchos casos, los problemas reales no son lineales.

PROBLEMAS

Ingeniería química/bioingeniería

28.1 Ejecute el primer cálculo de la sección 28.1, pero para el caso en que h = 10. Para obtener soluciones utilice los métodos de Heun (sin iteración) y de RK de cuarto orden.

28.2 Efectúe el segundo cálculo de la sección 28.1, pero para el sistema que se describe en el problema 12.4.

28.3 Un balance de masa para un producto químico completamente mezclado en un reactor se escribe así

$$V\frac{dc}{dt} = F - Qc - kVc^2$$

donde V = volumen (12 m^3), c = concentración (g/m^3), F = tasa de alimentación (175 g/min), Q = tasa de flujo (1 m^3/min) y k = tasa de reacción de segundo orden (0.15 m^3/g/min). Si $c(0) = 0$, resuelva la EDO hasta que la concentración alcance un nivel estable. Use el método del punto medio (h = 0.5) y grafique sus resultados.

Pregunta adicional: Si se ignora el hecho de que las concentraciones deben ser positivas, encuentre un rango de condiciones iniciales de modo que se obtenga una trayectoria muy diferente de la que se obtuvo con $c(0) = 0$. Relacione sus resultados con las soluciones de estado estable.

28.4 Si $c_{\text{ent}} = c_b(1 - e^{-0.12t})$, calcule la concentración en el flujo de salida de una sustancia conservativa (no reactiva) para un reactor único mezclado completamente, como función del tiempo. Use el método de Heun (sin iteración) para efectuar el cálculo. Emplee valores de c_b = 40 mg/m^3, Q = 6 m^3/min, V = 100 m^3 y c_0 = 20 mg/m^3. Haga el cálculo de t = 0 a 100 min con h = 2. Grafique sus resultados junto con la concentración del flujo de entrada *versus* el tiempo.

28.5 Se bombea agua de mar con una concentración de 8 000 g/m^3 hacia un tanque bien mezclado, a una tasa de 0.6 m^3/h. Debido al diseño defectuoso, el agua se evapora del tanque a una tasa de 0.025 m^3/h. La solución salina abandona el tanque a una tasa de 0.6 m^3/h.

a) Si originalmente el tanque contiene 1 m^3 de la solución que entra, ¿cuánto tiempo después de que se enciende la bomba de salida quedará seco el tanque?

b) Use métodos numéricos para determinar la concentración de sal en el tanque como función del tiempo.

28.6 Un cubo de hielo esférico (una "esfera de hielo") que mide 6 cm de diámetro es retirada de un congelador a 0°C y colocada en una pantalla de malla a temperatura ambiente T_a = 20°C. ¿Cuál será el

diámetro del cubo de hielo como función del tiempo fuera del congelador (si se supone que toda el agua que se funde gotea de inmediato a través de la pantalla)? El coeficiente de transferencia de calor h para una esfera en un cuarto tranquilo es alrededor de 3 W/(m^2 · K). El flujo calorífico de la esfera de hielo al aire está dado por

$$\text{Flujo} = \frac{q}{A} = h(T_a - T)$$

donde q = calor y A = área superficial de la esfera. Use un método numérico para hacer el cálculo. Observe que el calor latente de la fusión es de 333 kJ/kg, y la densidad del hielo es aproximadamente de 0.917 kg/m^3.

28.7 Las ecuaciones siguientes definen la concentración de tres reactivos:

$$\frac{dc_a}{dt} = -10c_a c_c + c_b$$

$$\frac{dc_b}{dt} = 10c_a c_c - c_b$$

$$\frac{dc_c}{dt} = -10c_a c_c + c_b - 2c_c$$

Si las condiciones iniciales son de $c_a = 50$, $c_b = 0$ y $c_c = 40$, encuentre las concentraciones para los tiempos de 0 a 3 s.

28.8 El compuesto A se difunde a través de un tubo de 4 cm de largo y reacciona conforme se difunde. La ecuación que gobierna la difusión con la reacción es

$$D\frac{d^2A}{dx^2} - kA = 0$$

En un extremo del tubo se encuentra una fuente grande de A con concentración de 0.1 M. En el otro extremo del tubo está un material que absorbe con rapidez cualquier A y hace que la concentración sea 0 M. Si $D = 1.5 \times 10^{-6}$ cm^2/s y $k = 5 \times 10^{-6}$s^{-1}, ¿cuál es la concentración de A como función de la distancia en el tubo?

28.9 En la investigación de un homicidio o de una muerte accidental, con frecuencia es importante estimar el tiempo que ha transcurrido desde la muerte. De observaciones experimentales, se sabe que la temperatura superficial de un objeto cambia con una tasa proporcional a la diferencia entre la temperatura del objeto y la del ambiente circundante, o temperatura ambiente. Esto se conoce como ley de Newton del enfriamiento. Así, si $T(t)$ es la temperatura del objeto al tiempo t, y T_a es la temperatura ambiente constante:

$$\frac{dT}{dt} = -K(T - T_a)$$

donde $K > 0$ es una constante de proporcionalidad. Suponga que en el momento $t = 0$ se descubre un cuerpo y se mide su temperatura, T_0. Se supone que en el momento de la muerte, la temperatura del cuerpo, T_d, era el valor normal de 37°C. Suponga que la temperatura del cuerpo al ser descubierto era de 29.5°C, y que dos horas después era de 23.5°C. La temperatura ambiente es de 20°C.

a) Determine K y el tiempo de la muerte.

b) Resuelva la EDO en forma numérica y grafique los resultados.

28.10 La reacción $A \to B$ tiene lugar en dos reactores en serie. Los reactores están bien mezclados pero no en estado estable. El balance de masa de estado no estable para cada tanque de agitado de los reactores es el siguiente:

$$\frac{dCA_1}{dt} = \frac{1}{\tau}(CA_0 - CA_1) - kCA_1$$

$$\frac{dCB_1}{dt} = -\frac{1}{\tau}CB_1 + kCA_1$$

$$\frac{dCA_2}{dt} = \frac{1}{\tau}(CA_1 - CA_2) - kCA_2$$

$$\frac{dCB_2}{dt} = \frac{1}{\tau}(CB_1 - CB_2) - kCA_2$$

donde CA_0 = concentración de A en la entrada del primer reactor, CA_1 = concentración de A a la salida del primer reactor (y en la entrada del segundo), CA_2 = concentración de A en la salida del segundo reactor. CB_1 = concentración de B en la salida del primer reactor (y en la entrada del segundo), CB_2 = concentración de B en el segundo reactor, τ = tiempo de residencia de cada reactor y k = tasa constante para la reacción de A para producir B. Si CA_0 es igual a 20, encuentre las concentraciones de A y B en ambos reactores durante sus primeros 10 minutos de operación. Utilice $k = 0.12$/min y $\tau = 5$ min, y suponga que las condiciones iniciales de todas las variables dependientes son cero.

28.11 Un reactor de procesamiento por lotes no isotérmico está descrito por las ecuaciones siguientes:

$$\frac{dC}{dt} = -e^{(-10/(T+273))}C$$

$$\frac{dT}{dt} = 1\,000e^{(-10/(T+273))}C - 10(T - 20)$$

donde C es la concentración del reactante y T es la temperatura del reactor. Inicialmente, el reactor se encuentra a 15°C y tiene una concentración de reactante C de 1.0 gmol/L. Encuentre la concentración y temperatura del reactor como función del tiempo.

28.12 El sistema siguiente es un ejemplo clásico de EDO rígidas que ocurre en la solución de una reacción química cinética:

$$\frac{dc_1}{dt} = -0.013c_1 - 1\,000c_1c_3$$

$$\frac{dc_2}{dt} = -2\,500c_2c_3$$

$$\frac{dc_3}{dt} = -0.013c_1 - 1\,000c_1c_3 - 2\,500c_2c_3$$

Resuelva las ecuaciones de $t = 0$ a 50, con condiciones iniciales $c_1(0) = c_2(0) = 1$ y $c_3(0) = 0$. Si usted tiene acceso al software de MATLAB, use tanto la función estándar (por ejemplo, `ode45`) como la rígida (por ejemplo, `ode23s`) para obtener sus soluciones.

28.13 Una *biopelícula* con un espesor L_f [cm] se desarrolla en la superficie de un sólido (figura P28.13). Después de atravesar una capa de espesor L [cm], un compuesto químico A se difunde al interior de la biopelícula donde está sujeta a una reacción irreversible de primer orden que lo convierte en un producto B.

Se pueden usar balances de masa de estado estable para deducir las siguientes ecuaciones diferenciales ordinarias para el compuesto A:

$$D\frac{d^2c_a}{dx^2} = 0 \qquad 0 \le x < L$$

$$D_f\frac{d^2c_a}{dx^2} - kc_a = 0 \qquad L \le x < L + L_f$$

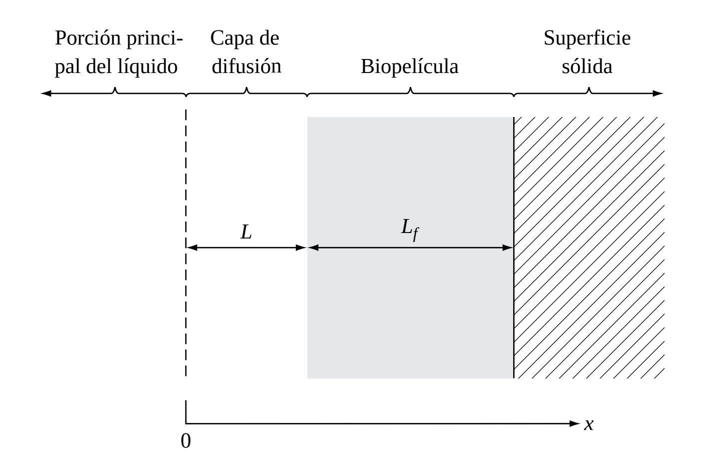


Figura P28.13
Una biopelícula desarrollándose sobre una superficie sólida.

donde D = coeficiente de difusión en la capa de difusión = 0.8 cm²/d, D_f = coeficiente de difusión de la biopelícula = 0.64 cm²/d y k = rapidez de primer orden para la conversión de A a B = 0.1/d. Se mantienen las siguientes condiciones de frontera:

$$c_a = c_{a0} \qquad \text{en } x = 0$$

$$\frac{dc_a}{dx} = 0 \qquad \text{en } x = L + L_f$$

donde c_{a0} = concentración de A en la porción principal del líquido = 100 mol/L. Use el método de diferencia finita para calcular la distribución en estado estable de A desde $x = 0$ hasta $L + L_f$, donde L = 0.008 cm y L_f = 0.004 cm. Utilice diferencias finitas centradas con Δx = 0.001cm.

28.14 La siguiente ecuación diferencial describe la concentración en estado estable de una sustancia que reacciona con cinética de primer orden en un reactor de flujo pistón con dispersión axial (figura P28.14).

$$D\frac{d^2c}{dx^2} - U\frac{dc}{dx} - kc = 0$$

donde D = coeficiente de dispersión [m²/h], c = concentración [mol/L], x = distancia [m], U = velocidad diagonal [m/h] y k = rapidez de reacción [/h]. Las condiciones de frontera se pueden formular como

$$Uc_{\text{ent}} = Uc(x = 0) - D\frac{dc}{dx}(x = 0)$$

$$\frac{dc}{dx}(x = L) = 0$$

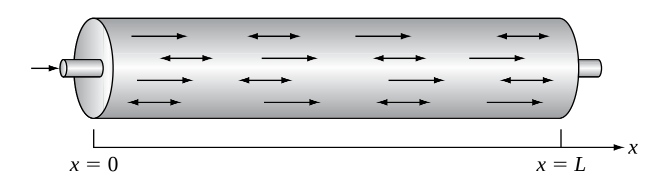


Figura P28.14
Un reactor de flujo pistón con dispersión axial.

donde c_{ent} = concentración en flujo de entrada [mol/L] y L = longitud del reactor [m]. Estas condiciones se llaman *condiciones de frontera Danckwerts*. Use el método de diferencias finitas para despejar la concentración como función de la distancia dados los siguientes parámetros: D = 5 000 m²/h, U = 100 m/h, k = 2/h, L = 100 m y c_{ent} = 100 mol/L. Utilice aproximaciones por diferencias finitas centradas con Δx = 10 m para obtener sus soluciones. Compare sus resultados numéricos con la solución analítica,

$$c = \frac{Uc_{ent}}{(U - D\lambda_1)\lambda_2 e^{\lambda_2 L} - (U - D\lambda_2)\lambda_1 e^{\lambda_1 L}} \times \left(\lambda_2 e^{\lambda_2 L} e^{\lambda_1 x} - \lambda_1 e^{\lambda_1 L} e^{\lambda_2 x}\right)$$

donde

$$\frac{\lambda_1}{\lambda_2} = \frac{U}{2D}\left(1 \pm \sqrt{1 + \frac{4kD}{U^2}}\right)$$

28.15 Una serie de reacciones de primer orden en fase líquida genera un producto deseable (B) y un producto indeseable (C):

$$A \xrightarrow{k_1} B \xrightarrow{k_2} C$$

Si las reacciones tienen lugar en un reactor de flujo pistón con dispersión axial (figura P28.14), se pueden usar balances de masa de estado estable para desarrollar las siguientes EDO de segundo orden:

$$D\frac{d^2c_a}{dx^2} - U\frac{dc_a}{dx} - k_1c_a = 0$$

$$D\frac{d^2c_b}{dx^2} - U\frac{dc_b}{dx} + k_1c_a - k_2c_b = 0$$

$$D\frac{d^2c_c}{dx^2} - U\frac{dc_c}{dx} + k_2c_b = 0$$

Use el método de diferencias finitas para despejar la concentración de cada reactante como función de la distancia, dados: D = 0.1 m²/min, U = 1 m/min, k_1 = 3/min, k_2 = 1/min, L = 0.5 m, $c_{a,\text{ent}}$ = 10 mol/L. Utilice aproximaciones por diferencias finitas centradas con Δx = 0.05 m para obtener sus soluciones y suponga condiciones de frontera Danckwerts como se describen en el problema 28.14. También calcule la suma de los reactantes como función de la distancia. ¿Tienen sentido sus resultados?

28.16 En un reactor por cargas se desarrollan bacterias que utilizan una fuente soluble de alimentación (sustrato) como se describe en la figura P28.16. La ingestión del sustrato se representa por un modelo logístico con limitación Michaelis-Menten. La muerte de las bacterias produce detritus que luego se convierte en sustrato por hidrólisis. Además, las bacterias también excretan directamente algo de sustra-

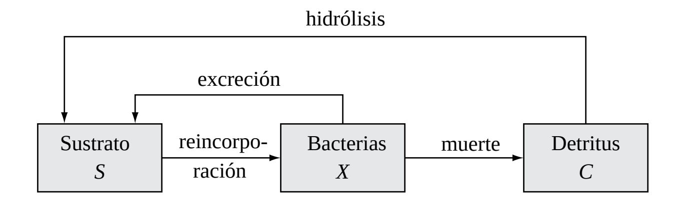


Figura P28.16

to. La muerte, la hidrólisis y la excreción se simulan como reacciones de primer orden.

Los balances de masa se pueden escribir como

$$\frac{dX}{dt} = \mu_{\text{máx}}\left(1 - \frac{X}{K}\right)\left(\frac{S}{K_s + S}\right) X - k_d X - k_e X$$
$$\frac{dC}{dt} = k_d X - k_h C$$
$$\frac{dS}{dt} = k_e X + k_h C - \mu_{\text{máx}}\left(1 - \frac{X}{K}\right)\left(\frac{S}{K_s + S}\right) X$$

donde X, C y S = concentraciones [mg/L] de bacterias, detritus y sustrato, respectivamente; $\mu_{\text{máx}}$ = máxima tasa de crecimiento [/d], K = capacidad logística de transporte [mg/L]; K_s = constante de media saturación Michaelis-Menten [mg/L], k_d = tasa de mortalidad [/d]; k_e = tasa de excreción [/d] y k_h = tasa de hidrólisis [/d]. Simule las concentraciones desde t = 0 hasta 100 d, dadas las condiciones iniciales $X(0)$ = 1 mg/L, $S(0)$ = 100 mg/L y $C(0)$ = 0 mg/L. Utilice los siguientes parámetros en su cálculo: $\mu_{\text{máx}}$ = 10/d, K = 10 mg/L, K_s = 10 mg/L, k_d = 0.1/d, k_e = 0.1/d y k_h = 0.1/d.

Ingeniería civil/ambiental

28.17 Ejecute el mismo cálculo para el sistema de Lotka-Volterra de la sección 28.2, pero utilice el método de *a*) Euler, *b*) Heun (sin iterar el corrector), *c*) RK de cuarto orden y *d*) la función `ode45` de MATLAB. En todos los casos use variables de precisión sencilla, tamaño de paso de 0.1, y simule de t = 0 a 20. Elabore gráficas de espacio-fase para todos los casos.

28.18 Ejecute el mismo cálculo para las ecuaciones de Lorenz de la sección 28.2, pero use el método de *a*) Euler, *b*) Heun (sin iterar el corrector), *c*) RK de cuarto orden y *d*) la función `ode45` de MATLAB. En todos los casos emplee variables de precisión sencilla y un tamaño de paso de 0.1 y simule de t = 0 a 20. Para todos los casos desarrolle gráficas de espacio-fase.

28.19 La ecuación siguiente se utiliza para modelar la deflexión del mástil de un bote sujeto a la fuerza del viento:

$$\frac{d^2y}{dz^2} = \frac{f}{2EI}(L - z)^2$$

donde f = fuerza del viento, E = módulo de elasticidad, L = longitud del mástil e I = momento de inercia. Calcule la deflexión si y = 0 y dy/dz = 0 en z = 0. Para su cálculo utilice valores de parámetro de f = 60, L = 30, E = 1.25 × 10^8 e I = 0.05.

28.20 Efectúe el mismo cálculo que en el problema 28.19, pero en vez de usar una fuerza del viento constante, emplee una fuerza que varíe con la altura de acuerdo con la ecuación (recuerde la sección 24.2)

$$f(z) = \frac{200z}{5 + z} e^{-2z/30}$$

28.21 Un ingeniero ambiental está interesado en estimar la mezcla que ocurre entre un lago estratificado y una bahía adyacente (véase la figura P28.21). Un trazador conservativo se mezcla instantáneamente con el agua de la bahía y después se monitorea la concentración del trazador durante el periodo que se muestra a continuación en los tres segmentos. Los valores son

t	0	2	4	6	8	12	16	20
c_1	0	15	11	7	6	3	2	1
c_2	0	3	5	7	7	6	4	2
c_3	100	48	26	16	10	4	3	2

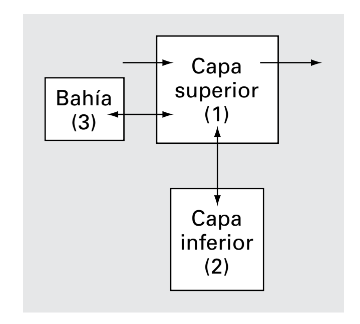


Figura P28.21

Con el empleo de balances de masa, el sistema puede modelarse con las EDO simultáneas siguientes:

$$V_1 \frac{dc_1}{dt} = -Qc_1 + E_{12}(c_2 - c_1) + E_{13}(c_3 - c_1)$$
$$V_2 \frac{dc_2}{dt} = E_{12}(c_1 - c_2)$$
$$V_3 \frac{dc_3}{dt} = E_{13}(c_1 - c_3)$$

donde V_i = volumen del segmento i, Q = flujo y E_{ij} = la tasa de mezcla difusiva entre los segmentos i y j. Utilice los datos y las ecuaciones diferenciales para estimar las E si V_1 = 1 × 10^7, V_2 = 8 × 10^6, V_3 = 5 × 10^6 y Q = 4 × 10^6. Para su análisis, emplee el método de Euler con tamaño de paso de 0.1.

28.22 Las dinámicas del crecimiento de la población son importantes en varios estudios de planeación como el transporte y la ingeniería de los recursos hidráulicos. Uno de los modelos más simples de dicho crecimiento incorpora el supuesto de que la tasa de cambio de la población p es proporcional a la que existe en cualquier momento t:

$$\frac{dp}{dt} = Gp \tag{P28.22.1}$$

donde G = tasa de crecimiento (anual). Este modelo tiene sentido intuitivo porque entre mayor sea la población más grande será el número de padres potenciales. Al tiempo t = 0, una isla tiene una población de 6 000 personas. Si G = 0.075 por año, emplee el método de Heun (sin iteración) para predecir la población en t = 20 años, con el uso de un tamaño de paso de 0.5 años. Grafique *p versus t*, en forma estándar y en escala semilogarítmico. Determine la pendiente de la línea sobre la gráfica semilogarítmica. Analice sus resultados.

28.23 Aunque el modelo del problema 28.22 funciona en forma adecuada cuando el crecimiento de la población es ilimitado, falla ante la existencia de factores como falta de comida, contaminación y falta de espacio, los cuales inhiben el crecimiento. En tales casos, la tasa de crecimiento se considera que es inversamente proporcional a la población. Un modelo de esta relación es

$$G = G'(p_{\text{máx}} - p) \tag{P28.23.1}$$

donde G' = tasa de crecimiento dependiente de la población (por persona-año) y $p_{\text{máx}}$ = población máxima sustentable. Así, cuando la población es pequeña ($p \ll p_{\text{máx}}$), la tasa de crecimiento será elevada y constante de $G' p_{\text{máx}}$. En tales casos, el crecimiento es ilimitado y la ecuación (P28.23.1) en esencia es idéntica a la (P28.22.1). Sin

embargo, conforme la población crece (es decir, conforme p se aproxima a $p_{máx}$), G disminuye hasta cero, esto es cuando $p = p_{máx}$. Así, el modelo predice que, cuando la población alcanza el nivel máximo sostenible, el crecimiento es inexistente, y el sistema se encontrará en estado estable. Al sustituir la ecuación (P28.23.1) en la (P28.22.1) se llega a

$$\frac{dp}{dt} = G'(p_{máx} - p)p$$

Para la misma isla que se estudió en el problema 28.22, emplee el método de Heun (sin iteración) para predecir la población en $t = 20$ años, con el uso de un tamaño de paso de 0.5 años. Emplee valores de $G = 10^{-5}$ por persona-año y $p_{máx} = 20\,000$ personas. Al tiempo $t = 0$, la isla tiene una población de 6 000 personas. Grafique *p versus t* e interprete la forma de la curva.

28.24 El Parque Nacional Isla Royal es un archipiélago de 210 millas cuadradas compuesto de una sola isla grande y muchas pequeñas, en el lago Superior. Alrededor de 1900 llegaron alces y hacia 1930, su población se acercaba a 3 000, por lo que devastaban la vegetación. En 1949, los lobos cruzaron un puente de hielo desde Ontario. Desde finales de la década de 1950, se registran los números de alces y lobos, como se muestra a continuación. (Un guión indica que no hay datos.)

Año	Alces	Lobos	Año	Alces	Lobos
1960	700	22	1972	836	23
1961	—	22	1973	802	24
1962	—	23	1974	815	30
1963	—	20	1975	778	41
1964	—	25	1976	641	43
1965	—	28	1977	507	33
1966	881	24	1978	543	40
1967	—	22	1979	675	42
1968	1 000	22	1980	577	50
1969	1 150	17	1981	570	30
1970	966	18	1982	590	13
1971	674	20	1983	811	23

a) Integre las ecuaciones de Lotka-Volterra de 1960 a 2020. Determine los valores de los coeficientes que arrojan un ajuste óptimo. Compare su simulación con los datos que usan un enfoque de series de tiempo y comente los resultados.

b) Grafique la simulación de *a*), pero emplee un enfoque de espacio-fase.

c) Después de 1993, suponga que los administradores de la vida silvestre atrapan un lobo por año y lo llevan fuera de la isla. Pronostique cómo evolucionaría tanto la población de lobos como de alces hacia el año 2020. Presente sus resultados tanto como una serie de tiempo como una gráfica de espacio-fase. Para este caso, así como para el inciso *d*), use los coeficientes que siguen: $a = 0.3$, $b = 0.01111$, $c = 0.2106$, $d = 0.0002632$.

d) Suponga que en 1993, algunos cazadores furtivos incursionaron en la isla y mataron al 50% de los alces. Pronostique cómo evolucionaría la población tanto de lobos como de alces hacia el año 2020. Presente sus resultados en gráficas tanto de series de tiempo como de espacio-fase.

28.25 Un cable cuelga de dos apoyos en *A* y *B* (véase la figura P28.25). El cable sostiene una carga distribuida cuya magnitud varía con *x* según la ecuación

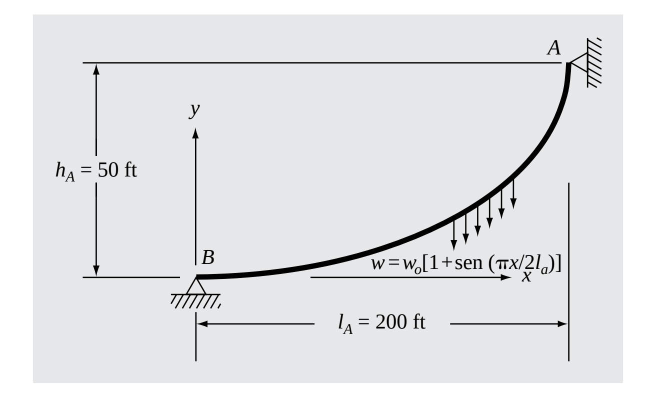


Figura P28.25

$$w = w_o\left[1 + \text{sen}\left(\frac{\pi x}{2l_A}\right)\right]$$

donde w_0 = 1 000 lbs/ft. La pendiente del cable $(dy/dx) = 0$ en $x = 0$, que es el punto más bajo del cable. También es el punto donde la tensión del cable alcanza un mínimo de T_0. La ecuación diferencial que gobierna el cable es

$$\frac{d^2y}{dx^2} = \frac{w_o}{T_o}\left[1 + \text{sen}\left(\frac{\pi x}{2l_A}\right)\right]$$

Resuelva esta ecuación con el uso de un método numérico y grafique la forma del cable (*y versus x*). Para la solución numérica, se desconoce el valor de T_0, por lo que la solución debe utilizar una técnica iterativa, similar al método del disparo, para converger en un valor correcto de h_A para distintos valores de T_0.

28.26 La ecuación diferencial básica de la curva elástica para una viga volada (véase la figura P28.26) está dada por

$$EI\frac{d^2y}{dx^2} = -P(L - x)$$

donde E = módulo de elasticidad e I = momento de inercia. Resuelva para la deflexión de la viga con el empleo de un método numérico. Se aplican los valores siguientes de parámetro: $E = 30\,000$ ksi, $I = 800$ in^4, $P = 1$ kip, $L = 10$ ft. Compare sus resultados numéricos con la solución analítica,

$$y = -\frac{PLx^2}{2EI} + \frac{Px^3}{6EI}$$

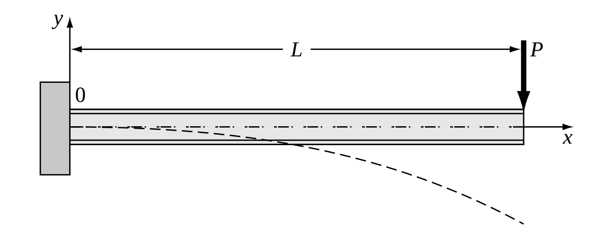


Figura P28.26

28.27 La ecuación diferencial básica de la curva elástica para una viga con carga uniforme (véase la figura P28.27) está dada por

$$EI = \frac{d^2y}{dx^2} = \frac{wLx}{2} - \frac{wx^2}{2}$$

donde E = módulo de elasticidad e I = momento de inercia. Resuelva para la deflexión de la viga usando *a*) la aproximación de diferencias finitas ($\Delta x = 2$ ft) y *b*) el método de disparo. Aplique los siguientes valores de parámetros: $E = 30\,000$ ksi, $I = 800$ in^4, $w = 1$ kip/ft, $L = 10$ ft. Compare sus resultados numéricos con la solución analítica,

$$y = \frac{wLx^3}{12EI} - \frac{wx^4}{24EI} - \frac{wL^3x}{24EI}$$

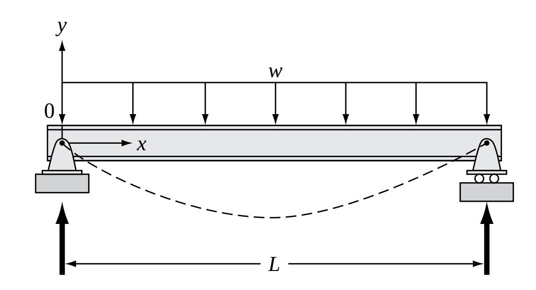


Figura P28.27

28.28 Un estanque se drena a través de un tubo como se observa en la figura P28.28. Con suposiciones simplificadoras, la ecuación diferencial siguiente describe cómo cambia la profundidad con el tiempo:

$$\frac{dh}{dt} = -\frac{\pi d^2}{4A(h)}\sqrt{2g(h+e)}$$

donde h = profundidad (m), t = tiempo (s), d = diámetro del tubo (m), $A(h)$ = área de la superficie del estanque como función de la profundidad (m^2), g = constante gravitacional (= 9.81 m/s^2) y e = profundidad de la salida del tubo por debajo del fondo del estanque (m). Con base en la tabla siguiente de área-profundidad, resuelva esta ecuación diferencial para determinar cuánto tiempo tomaría que el estanque se vaciara dado que $h(0) = 6$ m, $d = 0.25$ m, $e = 1$ m.

h, m	6	5	4	3	2	1	0
A(*h*), 10^4 m^2	1.17	0.97	0.67	0.45	0.32	0.18	0

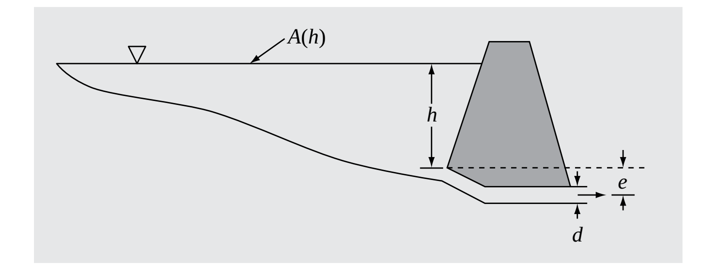


Figura P28.28

28.29 Los ingenieros y científicos utilizan modelos masa-resorte para entender la dinámica de las estructuras sujetas a la influencia de disturbios, como terremotos. En la figura P28.29 se ilustra una representación como ésas para un edificio de tres plantas. En este caso, el análisis se limita al movimiento horizontal de la estructura. Los balances de fuerza que se desarrollan para este sistema son los siguientes

$$\left(\frac{k_1+k_2}{m_1} - \omega^2\right)X_1 \quad - \frac{k_2}{m_1} \quad X_2 \quad = 0$$

$$-\frac{k_2}{m_2} \quad X_1 + \left(\frac{k_2+k_3}{m_2} - \omega^2\right)X_2 - \frac{k_3}{m_2}X_3 = 0$$

$$-\frac{k_3}{m_3} \quad X_2 + \left(\frac{k_3}{m_3} - \omega^2\right)X_3 = 0$$

Determine los valores y vectores propios y represente en forma gráfica los modos de vibración de la estructura por medio de dibujar las amplitudes *versus* la altura para cada uno de los vectores propios. Normalice las amplitudes de modo que el desplazamiento del tercer piso sea igual a uno.

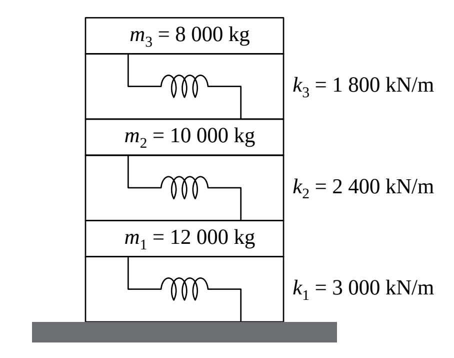


Figura P28.29

28.30 Con varias suposiciones de simplificación, la altura de estado estable del nivel hidrostático en un acuífero unidimensional no confinado (figura P28.30) se puede modelar con la siguiente EDO de segundo orden:

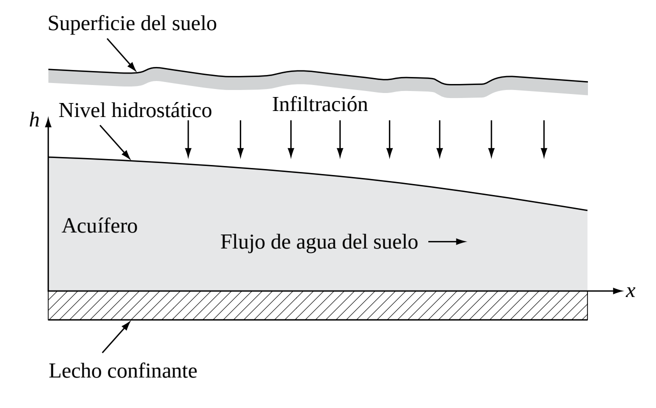


Figura P28.30
Un acuífero no confinado o "freático".

$$K\bar{h}\frac{d^2h}{dx^2} + N = 0$$

donde x = distancia (m), K = conductividad hidráulica [m/d], h = altura del nivel hidrostático [m], $\bar{h}$ = altura promedio del nivel hidrostático [m] y N = tasa de infiltración [m/d].

Despeje la altura del nivel hidrostático para $x = 0$ a 1 000 m donde $h(0) = 10$ m y $h(1\,000) = 5$ m. Use los siguientes parámetros para el cálculo: $K = 1$ m/d y $N = 0.0001$ m/d. Ponga la altura promedio del nivel hidrostático como el promedio de las condiciones de frontera. Obtenga su solución con *a*) el método de disparo y *b*) el método de diferencias finitas ($\Delta x = 100$ m).

28.31 En el problema 28.30 se usó un modelo linealizado de agua freática para simular la altura del nivel hidrostático para un acuífero no confinado. Se puede obtener un resultado más realista usando la siguiente EDO no lineal:

$$\frac{d}{dx}\left(Kh\frac{dh}{dx}\right) + N = 0$$

donde x = distancia [m], K = conductividad hidráulica [m/d], h = altura del nivel hidrostático [m] y N = tasa de infiltración [m/d]. Despeje la altura del nivel hidrostático para el mismo caso que en el problema 28.30. Es decir, despeje de $x = 0$ a 1 000 m con $h(0) = 10$ m y $h(1\,000) = 5$ m, $K = 1$ m/d y $N = 0.0001$ m/d. Obtenga su solución con *a*) el método del disparo y *b*) el método de diferencias finitas ($\Delta x = 100$ m).

28.32 Las ecuaciones de Lotka-Volterra que se describen en la sección 28.2 se han refinado para incluir factores adicionales que impactan la dinámica depredador-presa. Por ejemplo, además de la depredación, la población de presas puede estar limitada por otros factores como el espacio. Las limitaciones de espacio se pueden incorporar en el modelo como capacidad de acarreo (recuerde el modelo de logística descrito en el problema 28.16), como en

$$\frac{dx}{dt} = a\left(1 - \frac{x}{K}\right)x - bxy$$

$$\frac{dy}{dt} = -cy + dxy$$

donde K = capacidad de acarreo. Use los mismos valores de parámetros y condiciones iniciales que en la sección 28.2 para integrar estas ecuaciones de $t = 0$ a 100, usando `ode45`.

a) Use un valor muy alto de $K = 10^8$ para validar que usted obtiene los mismos resultados que en la sección 28.2.

b) Compare *a*) con una capacidad de acarreo más realista de $K = 200$. Explique sus resultados.

28.33 El crecimiento de algas unicelulares flotantes debajo de la descarga de una planta de tratamiento de aguas de drenaje se puede modelar con las siguientes EDO simultáneas:

$$\frac{da}{dt} = \left[k_g(n, p) - k_d - k_s\right]a$$

$$\frac{dn}{dt} = r_{nc}k_hc - r_{na}k_g(n, p)a$$

$$\frac{dp}{dt} = r_{pc}k_hc - r_{pa}k_g(n, p)a$$

$$\frac{dc}{dt} = r_{ca}k_da - k_hc$$

donde t = tiempo de viaje [d], a = concentración de clorofila en las algas [μgA/L], n = concentración de nitrógeno inorgánico [μgN/L], p = concentración de fósforo inorgánico [μgP/L], c = concentración de detritus [μgC/L], k_d = tasa de mortandad de las algas [/d], k_s = tasa de asentamiento de las algas [/d], k_h = tasa de hidrólisis de detritus [/d], r_{nc} = relación nitrógeno a carbono [μgN/μgC], r_{pc} = relación fósforo a carbono [μgP/μgC], r_{na} = relación nitrógeno a clorofila [μgN/μgA], r_{pa} = relación fósforo a clorofila [μgP/μgA] y $k_g(n, p)$ = tasa de crecimiento de las algas [/d], que se puede calcular con

$$k_g(n, p) = k_g \min\left\{\frac{p}{k_{sp} + p}, \frac{n}{k_{sn} + n}\right\}$$

donde k_g = tasa de crecimiento de las algas con exceso de nivel de nutriente [/d], k_{sp} = constante de media saturación de fósforo [μgP/L] y k_{sn} = constante de media saturación de nitrógeno [μgN/L]. Use las funciones `ode45` y `ode15s` para resolver estas ecuaciones de $t = 0$ a 50 d, dadas las condiciones iniciales $a = 1$, $n = 4\,000$, $p = 800$ y $c = 0$. Observe que los parámetros son $k_d = 0.1$, $k_s = 0.15$, $k_h = 0.025$, $r_{nc} = 0.18$, $r_{pc} = 0.025$, $r_{na} = 7.2$, $r_{pa} = 1$, $r_{ca} = 40$, $k_g = 0.5$, $k_{sp} = 2$ y $k_{sn} = 15$. Desarrolle gráficas de ambas soluciones e interprete los resultados.

28.34 Se han propuesto las siguientes EDO como modelo de una epidemia:

$$\frac{dS}{dt} = -aSI$$

$$\frac{dI}{dt} = aSI - rI$$

$$\frac{dR}{dt} = rI$$

donde S = los individuos susceptibles, I = los infectados, R = los recuperados, a = la tasa de infección y r = la tasa de recuperación. Una ciudad tiene 10 000 habitantes, todos los cuales son susceptibles.

a) Si un solo individuo infectado entra a la ciudad en $t = 0$, calcule la progresión de la epidemia hasta que el número de individuos infectados caiga por debajo de 10. Use los siguientes parámetros: $a = 0.002$/(persona · semana) y $r = 0.15$/d. Desarrolle gráficas de serie de tiempo de todas las variables de estado. También genere una gráfica de plano de fase *S versus I versus R*.

b) Suponga que, después de la recuperación, hay una pérdida de inmunidad que causa que los individuos recuperados se vuelvan susceptibles. Este mecanismo de reinfección se puede calcular como ρR, donde ρ = la tasa de reinfección. Modifique el modelo para que incluya este mecanismo y repita los cálculos de *a*) usando $\rho = 0.015$/d.

Ingeniería eléctrica

28.35 Realice el mismo cálculo que en la primera parte de la sección 28.3, pero con $R = 0.025\ \Omega$.

28.36 Resuelva la EDO de la primera parte de la sección 8.3 de $t = 0$ a 0.5, con técnicas numéricas, si $q = 0.1$ e $i = -3.281515$ en $t = 0$. Utilice un valor de $R = 50$ y los mismos parámetros que en la sección 8.3.

28.37 Para un circuito sencillo RL, la ley de Kirchhoff del voltaje requiere que (si se cumple la ley de Ohm)

$$L\frac{di}{dt} + Ri = 0$$

donde i = corriente, L = inductancia y R = resistencia. Resuelva para i, si $L = 1$, $R = 1.5$ e $i(0) = 0.5$. Resuelva este problema en forma analítica y con algún método numérico. Presente sus resultados en forma gráfica.

28.38 En contraste con el problema 28.37, las resistencias reales no siempre siguen la ley de Ohm. Por ejemplo, la caída del voltaje quizá sea no lineal y la dinámica del circuito quede descrita por una relación como la siguiente

$$L\frac{di}{dt}+R\left[\frac{i}{I}-\left(\frac{i}{I}\right)^3\right]=0$$

donde todos los demás parámetros se definen como en el problema 28.37 e I es una corriente conocida de referencia e igual a 1. Resuelva para i como función del tiempo en las mismas condiciones que se especifican para el problema 28.37.

28.39 Desarrolle un problema de valor propio para una red LC similar a la de la figura 28.14, pero con sólo dos circuitos. Es decir, omita el circuito de i_3. Dibuje la red e ilustre la forma en que las corrientes oscilan en sus modos primarios.

28.40 Igual que se pueden usar la ley de Fourier y el balance térmico para caracterizar la distribución de temperaturas, hay disponibles relaciones análogas para modelar problemas de campo en otras áreas de la ingeniería. Por ejemplo, los ingenieros eléctricos usan un enfoque similar cuando modelan campos electrostáticos. Bajo un número de supuestos simplificadores, se puede representar un análogo de la ley de Fourier en forma unidimensional como

$$D=-\varepsilon\frac{dV}{dx}$$

donde D se llama vector de densidad del flujo eléctrico, ε = permitividad del material y V = potencial electrostático. De manera similar, una ecuación de Poisson para campos electrostáticos se puede representar en una dimensión como

$$\frac{d^2V}{dx^2}=-\frac{\rho_v}{\varepsilon}$$

donde ρ_v = densidad de carga. Use la técnica de diferencia finita con $\Delta x = 2$ para determinar V para un alambre en el que $V(0) = 1\,000$, $V(20) = 0$, $\varepsilon = 2$, $L = 20$ y $\rho_v = 30$.

Ingeniería mecánica/aeroespacial

28.41 Lleve a cabo el mismo cálculo que en la sección 28.4 pero para un péndulo de 1 m de longitud.

28.42 La tasa de enfriamiento de un cuerpo se expresa como

$$\frac{dT}{dt}=-k(T-T_a)$$

donde T = temperatura del cuerpo (°C), T_a = temperatura del medio circundante (°C) y k = constante de proporcionalidad (min^{-1}). Así, esta ecuación especifica que la tasa de enfriamiento es proporcional a la diferencia de la temperatura del cuerpo y del ambiente circundante. Si una bola de metal se calienta a 90°C y se sumerge en agua que se mantiene a un valor constante de T_a = 20°C, utilice un método numérico para calcular el tiempo que toma que la bola se enfríe a 40°C, si k = 0.25 min^{-1}.

28.43 La tasa de flujo calorífico (conducción) entre dos puntos de un cilindro calentado por un extremo está dada por

$$\frac{dQ}{dt}=\lambda A\frac{dT}{dx}$$

donde λ = una constante, A = área de la sección transversal del cilindro, Q = flujo calorífico, T = temperatura, t = tiempo y x = distancia a partir del extremo calentado. Debido a que la ecuación involucra dos derivadas, la ecuación se simplificará haciendo que

$$\frac{dT}{dx}=\frac{100(L-x)(20-t)}{100-xt}$$

donde L es la longitud de la barra. Combine las dos ecuaciones y calcule el flujo de calor de $t = 0$ a 25 s. La condición inicial es $Q(0) = 0$ y los parámetros son λ = 0.5 cal · cm/s, A = 12 cm², L = 20 cm y x = 2.5 cm. Grafique sus resultados.

28.44 Repita el problema del paracaidista (ejemplo 1.2), pero con la fuerza hacia arriba que se debe al arrastre igual a una tasa de segundo orden:

$$F_u=-cv^2$$

donde c = 0.225 kg/m. Resuelva de $t = 0$ a 30, grafique sus resultados y compárelos con los del ejemplo 1.2.

28.45 Imagine que después de caer durante 13 s, el paracaidista de los ejemplos 1.1 y 1.2, tira de la cuerda de apertura. En este punto, suponga que el coeficiente de arrastre se incrementa en forma instantánea a un valor constante de 55 kg/s. Calcule la velocidad del paracaidista de $t = 0$ a 30 s con el método de Heun (sin iteración del corrector) con un tamaño de paso de 2 s. Grafique L *versus* t, de $t = 0$ a 30 s.

28.46 La ecuación diferencial ordinaria siguiente describe el movimiento de un sistema amortiguado resorte-masa (véase la figura P28.46):

$$m\frac{d^2x}{dt^2}+a\left|\frac{dx}{dt}\right|\frac{dx}{dt}+bx^3=0$$

donde x = desplazamiento a partir de la posición de equilibrio, t = tiempo, m = 1 kg masa y a = 5 N/(m/s)². El término de amortiguamiento es no lineal y representa el amortiguamiento del aire.

El resorte es un resorte cúbico y también es no lineal con b = 5 N/m³. Las condiciones iniciales son

Velocidad inicial $\quad \frac{dx}{dt}=0.5\text{ m/s}$

Desplazamiento inicial $\quad x = 1$ m

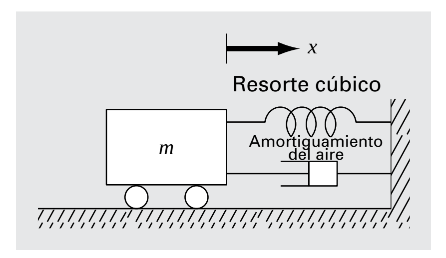


Figura P28.46

Resuelva esta ecuación con algún método numérico para el periodo de tiempo $0 \leq t \leq 8$ s. Grafique el desplazamiento y la velocidad *versus* el tiempo, y grafique el retrato espacio-fase (velocidad *versus* desplazamiento) para todos los casos siguientes:

a) Ecuación lineal similar

$$m\frac{d^2x}{dt^2}+2\frac{dx}{dt}+5x=0$$

b) La ecuación no lineal con sólo un término de resorte no lineal

$$\frac{d^2x}{dt^2}+2\frac{dx}{dt}+bx^3=0$$

c) La ecuación no lineal con sólo un término de amortiguamiento no lineal

$$m\frac{d^2x}{dt^2}+a\left|\frac{dx}{dt}\right|\frac{dx}{dt}+5x=0$$

d) La ecuación por completo no lineal en la que tanto el término de amortiguamiento como el de resorte son no lineales

$$m\frac{d^2x}{dt^2}+a\left|\frac{dx}{dt}\right|\frac{dx}{dt}+bx^3=0$$

28.47 Un sistema amortiguado y forzado resorte-masa (véase la figura P28.47) tiene la ecuación diferencial ordinaria siguiente para su movimiento:

$$m\frac{d^2x}{dt^2}+a\left|\frac{dx}{dt}\right|\frac{dx}{dt}+kx=F_o \text{ sen } (\omega t)$$

donde x = desplazamiento a partir de la posición de equilibrio, t = tiempo, m = 2 kg masa, a = 5 N/(m/s)^2 y k = 6 N/m. El término de amortiguamiento es no lineal y representa el amortiguamiento del aire. La función de fuerza $F_0 \text{sen}(\omega t)$ tiene valores de F_0 = 2.5 N y ω = 0.5 rad/s. Las condiciones iniciales son

Velocidad inicial	$\frac{dx}{dt} = 0\ m/s$
Desplazamiento inicial	$x = 1$ m

Resuelva esta ecuación con el empleo de algún método numérico durante el periodo de tiempo $0 \leq t \leq 15$ s. Grafique el desplazamiento y la velocidad *versus* el tiempo, y grafique la función de fuerza sobre la misma curva. Asimismo, desarrolle una gráfica separada de la velocidad *versus* el desplazamiento.

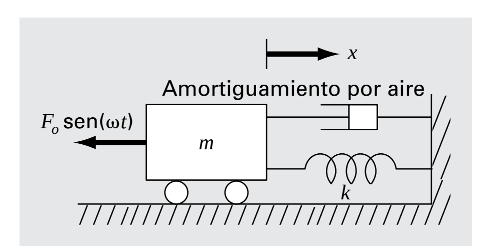


Figura P28.47

28.48 La distribución de temperatura en una aleta de enfriamiento cónica y ahusada (véase la figura P28.48) está descrita por la ecuación diferencial siguiente, que ha sido no dimensionada

$$\frac{d^2u}{dx^2}+\left(\frac{2}{x}\right)\left(\frac{du}{dx}-pu\right)=0$$

donde u = temperatura ($0 \leq u \leq 1$), x = distancia axial ($0 \leq x \leq 1$), y p es un parámetro no dimensional que describe la transferencia de calor y la geometría

$$p=\frac{hL}{k}\sqrt{1+\frac{4}{2m^2}}$$

donde h = coeficiente de transferencia de calor, k = conductividad térmica, L = longitud o altura del cono, y m = pendiente de la pared del cono. La ecuación tiene las condiciones de frontera siguientes

$$u(x=0)=0 \quad u(x=1)=1$$

Resuelva esta ecuación para la distribución de temperatura con el empleo de métodos de diferencias finitas. Para las derivadas utilice diferencias finitas exactas de segundo orden análogas. Escriba un programa de computadora para obtener la solución y grafique la temperatura *versus* la distancia axial para distintos valores de p = 10, 20, 50 y 100.

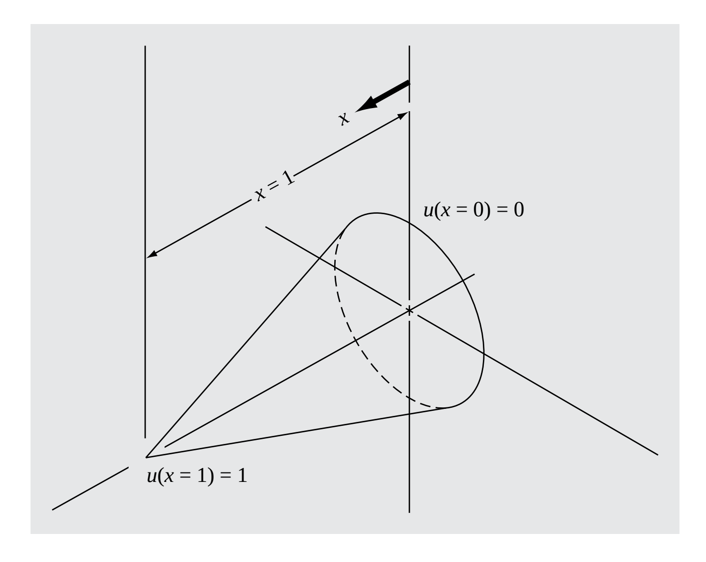


Figura P28.48

28.49 Las dinámicas de un sistema forzado resorte-masa- amortiguador se representa con la EDO de segundo orden siguiente:

$$m\frac{d^2x}{dt^2}+c\frac{dx}{dt}+k_1x+k_3x^3=P\cos(\omega t)$$

donde m = 1 kg, c = 0.4 N · s/m, P = 0.5 N y ω = 0.5/s. Utilice un método numérico para resolver cuál es el desplazamiento (x) y la velocidad ($L = dx/dt$) como función del tiempo con condiciones ini-

ciales $x = v = 0$. Exprese sus resultados en forma gráfica como gráficas de series de tiempo (x y L *versus* t) y gráfica del espacio-fase (v *versus* x). Haga simulaciones para un resorte *a*) lineal ($k_1 = 1$; $k_3 = 0$) y *b*) no lineal ($k_1 = 1$; $k_3 = 0.5$).

28.50 La ecuación diferencial para la velocidad de alguien que practica el salto del *bungee* es diferente según si el saltador ha caído una distancia en la que la cuerda está extendida por completo y comienza a encogerse. Así, si la distancia recorrida es menor que la longitud de la cuerda, el saltador sólo está sujeto a las fuerzas gravitacional y de arrastre. Una vez que la cuerda comienza a encogerse, también deben incluirse las fuerzas del resorte y del amortiguamiento de la cuerda. Estas dos condiciones se expresan con las ecuaciones siguientes:

$$\frac{dv}{dt} = g - \text{sign}(v)\frac{c_d}{m}v^2 \qquad x \le L$$

$$\frac{dv}{dt} = g - \text{sign}(v)\frac{c_d}{m}v^2 - \frac{k}{m}(x-L) - \frac{\gamma}{m}v \qquad x > L$$

donde v = velocidad (m/s), t = tiempo (s), g = constante gravitacional (= 9.81 m/s^2), sign(x) = función que devuelve –1, 0 y 1, para x negativa, cero y positiva, respectivamente, c_d = coeficiente de arrastre de segundo orden (kg/m), m = masa (kg), k = constante de elasticidad de la cuerda (N/m), γ = coeficiente de amortiguamiento de la cuerda (N · s/m) y L = longitud de la cuerda (m). Determine la posición y velocidad del saltador dadas por los parámetros siguientes: L = 30 m, m = 68.1 kg, c_d = 0.25 kg/m, k = 40 N/m y γ = 8 kg/s. Haga el cálculo de t = 0 a 50 s y suponga que las condiciones iniciales son $x(0) = v(0) = 0$.

28.51 Dos masas están fijadas a un muro mediante resortes lineales (figura P28.51). Los balances de fuerza basados en la segunda ley de Newton se pueden escribir como

$$\frac{d^2x_1}{dt^2} = -\frac{k_1}{m_1}(x_1 - L_1) + \frac{k_2}{m_1}(x_2 - x_1 - w_1 - L_2)$$

$$\frac{d^2x_2}{dt^2} = -\frac{k_2}{m_2}(x_2 - x_1 - w_1 - L_2)$$

donde k = constantes de elasticidad, m = masa, L = longitud del resorte sin estirar y ω = anchura de la masa. Calcule las posiciones de las masas como función del tiempo usando los siguientes valores de parámetros: $k_1 = k_2 = 5$, $m_1 = m_2 = 2$, $\omega_1 = \omega_2 = 5$ y $L_1 = L_2 = 2$. Establezca las condiciones iniciales como $x_1 = L_1$ y $x_2 = L_1 + w_1 + L_2 + 6$. Realice la simulación de t = 0 a 20. Construya gráficas de series de tiempo de ambos desplazamientos y las velocidades. Además, produzca una gráfica de plano de fase de x_1 *vs*. x_2.

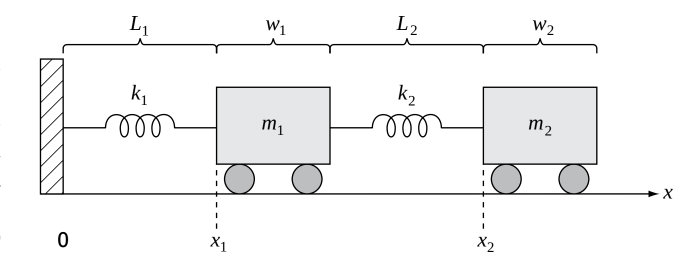


Figura P28.51

EPÍLOGO: PARTE SIETE

PT7.4 ALTERNATIVAS

La tabla PT7.3 muestra las ventajas y las desventajas de los métodos numéricos para la solución de ecuaciones diferenciales ordinarias con valor inicial. Los factores considerados en esta tabla deben ser analizados por el ingeniero cuando seleccione un método para aplicarse en cada problema específico.

Se pueden usar técnicas simples de autoinicio, como el método de Euler, si los requerimientos del problema presentan un intervalo corto de integración. En tal caso, es posible obtener una buena exactitud utilizando tamaños de paso pequeños para evitar grandes errores de truncamiento, y los errores de redondeo serán aceptables. El método de Euler también resulta apropiado en casos donde el modelo matemático tiene un alto nivel de incertidumbre, o tiene coeficientes o funciones de fuerza con errores significativos, como los que llegan a surgir en un proceso de medición. En este caso, la exactitud del modelo mismo simplemente no justifica el trabajo de cálculo requerido al emplear un método numérico más complicado. Por último, en ocasiones, las técnicas más simples son las mejores cuando el problema o la simulación necesitan realizarse sólo unas cuantas veces. En dichos problemas, quizá sea mejor usar un método simple que sea fácil de programar y de entender, a pesar de que el método pueda ser ineficiente en términos computacionales y relativamente lento al correrse en la computadora.

Si el intervalo de integración del problema es lo suficientemente grande como para necesitar un gran número de pasos, entonces puede ser necesario y adecuado emplear una técnica más exacta que el método de Euler. El método RK de cuarto orden es popular y confiable para muchos problemas de ingeniería. En tales casos, también se aconseja estimar el error de truncamiento en cada paso como una guía para seleccionar el mejor tamaño de paso. Esto se lleva a cabo con los procedimientos RK adaptativos o con el método de Adams de cuarto orden. Si los errores de truncamiento son muy pequeños, podría ser acertado aumentar el tamaño de paso para ahorrar tiempo de computadora. Por otro lado, si el error de truncamiento es grande, se deberá disminuir el tamaño de paso para

Tabla PT7.3 Comparación de las características de métodos alternativos para la solución numérica de EDO. Las comparaciones se basan en la experiencia general y no toman en cuenta el comportamiento de las funciones especiales.

Método	Valores iniciales	Iteraciones requeridas	Error global	Cambio de tamaño de paso	Dificultad de programación	Comentarios
Un paso						
De Euler	1	No	$O(h)$	Fácil	Escasa	Bueno para estimaciones rápidas
De Heun	1	Sí	$O(h^2)$	Fácil	Moderada	—
Punto medio	1	No	$O(h^2)$	Fácil	Moderada	—
Ralston de segundo orden	1	No	$O(h^2)$	Fácil	Moderada	El método RK de segundo orden que minimiza el error de truncamiento
RK de cuarto orden	1	No	$O(h^4)$	Fácil	Moderada	Ampliamente usado
Adaptativo de cuarto orden						
RK o RK-Fehlberg	1	No	$O(h^5)$*	Fácil	Moderada a difícil	La estimación del error permite ajuste del tamaño de paso
De pasos múltiples						
Heun sin autoinicio Heun	2	Sí	$O(h^3)$*	Difícil	Moderada a difícil†	Método de pasos múltiples simple
De Milne	4	Sí	$O(h^5)$*	Difícil	Moderada a difícil†	Algunas veces inestable
Adams de cuarto orden	4	Sí	$O(h^5)$*	Difícil	Moderada a difícil†	

* Siempre que la estimación del error se utilice para modificar la solución.
† Con tamaño de paso variable.

evitar la acumulación del error. Si se esperan problemas significativos de estabilidad, deberá evitarse el método de Milne. El método de Runge-Kutta es simple de programar y fácil de usar; aunque llega a ser menos eficiente que los métodos de pasos múltiples. Sin embargo, el método de Runge-Kutta a menudo se utiliza en cualquier caso para obtener los valores iniciales requeridos en los métodos de pasos múltiples.

Una gran cantidad de problemas de ingeniería pueden utilizar un intervalo intermedio de integración y pocos requerimientos de exactitud. En tales casos, los métodos RK de segundo orden y de Heun sin autoinicio resultan fáciles de usar, y son relativamente eficientes y exactos.

Los *sistemas rígidos* consideran ecuaciones con componentes que varían lenta y rápidamente. Por lo común, se requieren técnicas especiales para la solución adecuada de ecuaciones rígidas. Por ejemplo, se utilizan procedimientos implícitos. Usted puede consultar a Enright y cols. (1975), Gear (1971) y Shampine y Gear (1979) para obtener más información respecto a esas técnicas.

Existen varias técnicas para resolver problemas de valores propios. Para sistemas pequeños o cuando sólo se requieren unos pocos de los valores propios menores o mayores, es posible usar procedimientos simples como el método de polinomios o el de potencias. Para sistemas simétricos, se emplean los métodos de Jacobi, de Given o de Householder. Por último, el método QR representa un procedimiento general para encontrar todos los valores propios de matrices simétricas y no simétricas.

PT7.5 RELACIONES Y FÓRMULAS IMPORTANTES

La tabla PT7.4 resume la información importante que se presentó en la parte siete. Se recomienda consultar esta tabla para un rápido acceso a las relaciones y las fórmulas importantes.

PT7.6 MÉTODOS AVANZADOS Y REFERENCIAS ADICIONALES

Aunque hemos revisado varias técnicas para resolver ecuaciones diferenciales ordinarias, existe información adicional que es importante en la práctica de la ingeniería. El problema de la *estabilidad* se presentó en la sección 26.2.4. Este tema es de importancia relevante en todos los métodos para resolver EDO. Un análisis más amplio del tema se encuentra en Carnahan, Luther y Wilkes (1969), Gear (1971) y Hildebrand (1974).

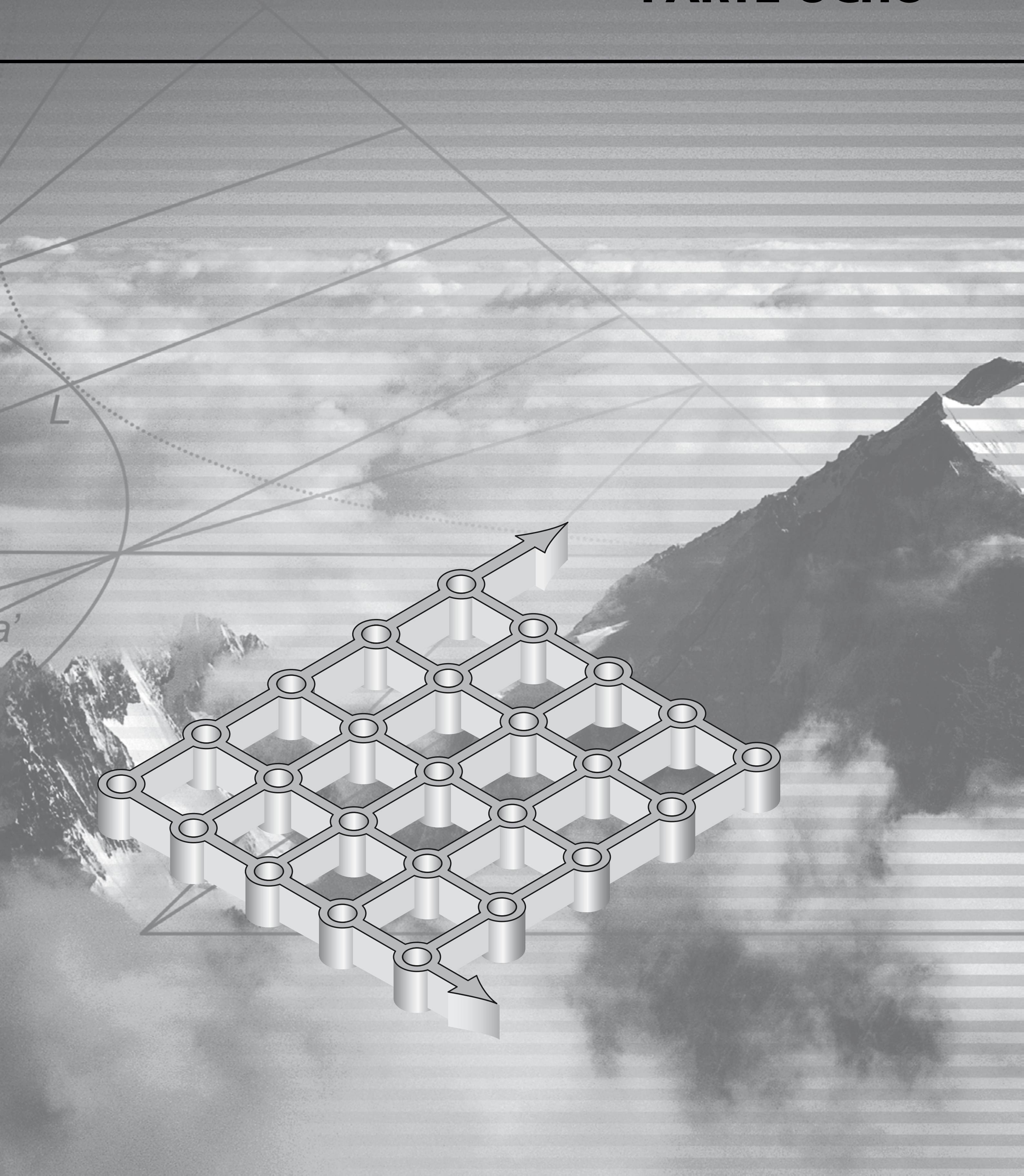
En el capítulo 27 presentamos los métodos para resolver problemas *con valores en la frontera*. Se sugiere consultar a Isaacson y Keller (1966), Keller (1968), Na (1979) y Scott y Watts (1976) para mayor información sobre problemas estándar con valores en la frontera. Material adicional acerca de los valores propios se encuentra en Ralston y Rabinowitz (1978), Wilkinson (1965), Fadeev y Fadeeva (1963), y Householder (1953, 1964).

En resumen, lo anterior pretende ofrecerle caminos para una exploración más profunda sobre el tema. Además, todas las referencias anteriores proporcionan descripciones de las técnicas básicas que se estudiaron en la parte siete. Le recomendamos consultar estas fuentes alternativas para ampliar su comprensión de los métodos numéricos en la solución de ecuaciones diferenciales.

Tabla PT7.4 Resumen de la información importante presentada en la parte siete.

Método	Formulación	Interpretación gráfica	Errores
Euler (RK de primer orden)	$y_{i+1} = y_i + hk_1$ $k_1 = f(x_i, y_i)$		Error local $\simeq O(h^2)$ Error global $\simeq O(h)$
RK de segundo orden de Ralston	$y_{i+1} = y_i + h(\frac{1}{3}k_1 + \frac{2}{3}k_2)$ $k_1 = f(x_i, y_i)$ $k_2 = f(x_i + \frac{3}{4}h, y_i + \frac{3}{4}hk_1)$		Error local $\simeq O(h^3)$ Error global $\simeq O(h^2)$
RK clásico de cuarto orden	$y_{i+1} = y_i + h(\frac{1}{6}k_1 + \frac{1}{3}k_2 + \frac{1}{3}k_3 + \frac{1}{6}k_4)$ $k_1 = f(x_i, y_i)$ $k_2 = f(x_i + \frac{1}{2}h, y_i + \frac{1}{2}hk_1)$ $k_3 = f(x_i + \frac{1}{2}h, y_i + \frac{1}{2}hk_2)$ $k_4 = f(x_i + h, y_i + hk_3)$		Error local $\simeq O(h^5)$ Error global $\simeq O(h^4)$
Heun sin autoinicio	Predictor (método de punto medio): $y_{i+1}^0 = y_{i-1}^m + 2hf(x_i, y_i^m)$		Modificador del predictor: $E_p \simeq \frac{4}{5}\left(y_{i,u}^m + y_{i,u}^0\right)$
	Corrector (regla del trapecio): $y_{i+1}^j = y_i^m + h\dfrac{f(x_i, y_i^m) + f(x_{i+1}, y_{i+1}^{j-1})}{2}$		Modificador del corrector: $E_c \simeq -\dfrac{y_{i+1,u}^m - y_{i+1,u}^0}{5}$
Adams de cuarto orden	Predictor (cuarto de Adams-Bashforth): $y_{i+1}^0 = y_i^m + h\left(\frac{55}{24}f_i^m - \frac{59}{24}f_{i-1}^m + \frac{37}{24}f_{i-2}^m - \frac{9}{24}f_{i-3}^m\right)$		Modificador del predictor: $E_p \simeq \frac{251}{270}\left(y_{i,u}^m - y_{i,u}^0\right)$
	Corrector (cuarto de Adams-Moulton): $y_{i+1}^j = y_i^m + h\left(\frac{9}{24}f_{i+1}^{j-1} + \frac{19}{24}f_i^m - \frac{5}{24}f_{i-1}^m + \frac{1}{24}f_{i-2}^m\right)$		Modificador del corrector: $E_c \simeq \frac{19}{270}\left(y_{i+1,u}^m - y_{i+1,u}^0\right)$

PARTE OCHO



ECUACIONES DIFERENCIALES PARCIALES

PT8.1 MOTIVACIÓN

Dada una función u que depende tanto de x como de y, la derivada parcial de u con respecto a x en un punto arbitrario (x, y) se define como

$$\frac{\partial u}{\partial x} = \lim_{\Delta x \to 0} \frac{u(x+\Delta x, y) - u(x, y)}{\Delta x} \tag{PT8.1}$$

De manera similar, la derivada parcial con respecto a y se define como

$$\frac{\partial u}{\partial y} = \lim_{\Delta y \to 0} \frac{u(x, y+\Delta y) - u(x, y)}{\Delta y} \tag{PT8.2}$$

Una ecuación que tiene derivadas parciales de una función desconocida, de dos o más variables independientes, se denomina *ecuación diferencial parcial*, o EDP. Por ejemplo,

$$\frac{\partial^2 u}{\partial x^2} + 2xy\frac{\partial^2 u}{\partial y^2} + u = 1 \tag{PT8.3}$$

$$\frac{\partial^3 u}{\partial x^2 \partial y} + x\frac{\partial^2 u}{\partial y^2} + 8u = 5y \tag{PT8.4}$$

$$\left(\frac{\partial^2 u}{\partial x^2}\right)^3 + 6\frac{\partial^3 u}{\partial x \partial y^2} = x \tag{PT8.5}$$

$$\frac{\partial^2 u}{\partial x^2} + xu\frac{\partial u}{\partial y} = x \tag{PT8.6}$$

El *orden* de una EDP es el de la derivada parcial de mayor orden que aparece en la ecuación. Por ejemplo, las ecuaciones (PT8.3) y (PT8.4) son de segundo y tercer orden, respectivamente.

Se dice que una ecuación diferencial parcial es *lineal*, si es lineal en la función desconocida y en todas sus derivadas, con coeficientes que dependen sólo de las variables independientes. Por ejemplo, las ecuaciones (PT8.3) y (PT8.4) son lineales; mientras que las ecuaciones (PT8.5) y (PT8.6) no lo son.

Debido a su amplia aplicación en ingeniería, nuestro estudio de las EDP se concentrará en las ecuaciones diferenciales lineales de segundo orden. Para dos variables independientes, tales ecuaciones se pueden expresar de la forma general siguiente:

$$A\frac{\partial^2 u}{\partial x^2} + B\frac{\partial^2 u}{\partial x \partial y} + C\frac{\partial^2 u}{\partial y^2} + D = 0 \tag{PT8.7}$$

donde A, B y C son funciones de x y y, y D es una función de x, y, u, $\partial u/\partial x$ y $\partial u/\partial y$. Dependiendo de los valores de los coeficientes de los términos de la segunda derivada (A, B y C), la ecuación (PT8.7) se clasifica en una de tres categorías (tabla PT8.1). Esta clasificación, que se basa en el método de las características (por ejemplo, véase Vichnevetsky, 1981, o Lapidus y Pinder, 1981), es útil debido a que cada categoría se relaciona con problemas de ingeniería específicos y distintos, que demandan

Tabla PT8.1 Categorías en las que se clasifican las ecuaciones diferenciales parciales lineales de segundo orden con dos variables.

$B^2 - 4AC$	Categoría	Ejemplo
< 0	Elíptica	Ecuación de Laplace (estado estacionario con dos dimensiones espaciales) $\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = 0$
= 0	Parabólica	Ecuación de conducción del calor (variable de tiempo y una dimensión espacial) $\frac{\partial T}{\partial t} = k' \frac{\partial^2 T}{\partial x^2}$
> 0	Hiperbólica	Ecuación de onda (variable de tiempo y una dimensión espacial) $\frac{\partial^2 y}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 y}{\partial t^2}$

técnicas de solución especiales. Deberá observarse que en los casos donde *A*, *B* y *C* dependen de *x* y *y*, la ecuación puede encontrarse en una categoría diferente, dependiendo de la ubicación en el dominio donde la ecuación se satisface. Por sencillez, limitaremos el presente análisis a las EDP que pertenecen exclusivamente a una de las categorías.

PT8.1.1 EDP y la práctica en ingeniería

Cada una de las categorías de ecuaciones diferenciales parciales en la tabla PT8.1 corresponde a una clase específica de problemas en ingeniería. Las secciones iniciales de los siguientes capítulos se dedicarán a obtener cada tipo de ecuación para un problema de ingeniería en particular. En principio, analizaremos sus propiedades generales y sus aplicaciones, y mostraremos cómo se emplean en diferentes contextos físicos.

Comúnmente, las *ecuaciones elípticas* se utilizan para caracterizar sistemas en *estado estacionario.* Como en la *ecuación de Laplace* de la tabla PT8.1, esto se indica por la ausencia de una derivada con respecto al tiempo. Así, estas ecuaciones se emplean para determinar la distribución en estado estacionario de una incógnita en dos dimensiones espaciales.

Un ejemplo sencillo es la placa caliente de la figura PT8.1*a*. En tal caso, los bordes de la placa se mantienen a temperaturas diferentes. Como el calor fluye de las regiones de alta temperatura a las de baja temperatura, las condiciones de frontera establecen un potencial que lleva el flujo de calor de la frontera caliente a la fría. Si transcurre suficiente tiempo, este sistema alcanzará al final la distribución de temperatura estable o en estado estacionario representada en la figura PT8.1*a*. La ecuación de Laplace, junto con las condiciones de frontera adecuadas, ofrece un medio para determinar esta distribución. Por analogía, se puede utilizar el mismo procedimiento para abordar otros problemas que implican potenciales, como la filtración de agua bajo una presa (figura PT8.1*b*) o la distribución de un campo eléctrico (figura PT8.1*c*).

A diferencia de la categoría elíptica, las *ecuaciones parabólicas* determinan cómo una incógnita varía tanto en el espacio como en el tiempo, lo cual se manifiesta por la presencia de las derivadas espacial y temporal, como la *ecuación de conducción de calor* considerada en la tabla PT8.1. Tales casos se conocen como *problemas de propagación*, puesto que la solución se “propaga”, o cambia, con el tiempo.

Un ejemplo sencillo es el de una barra larga y delgada aislada, excepto en sus extremos (figura PT8.2*a*). El aislamiento se emplea para evitar complicaciones debido a la pérdida de calor a lo largo de la barra. Como en el caso de la placa caliente de la figura PT8.1*a*, los extremos de la barra se encuentran a una temperatura fija. Sin embargo, a diferencia de la figura PT8.1*a*, el espesor de la barra nos permite suponer que el calor se distribuye de manera uniforme sobre su sección transversal (es decir, lateralmente). En consecuencia, el flujo de calor lateral no es un problema, y el problema se reduce a estudiar la conducción del calor a lo largo del eje longitudinal de la barra. En lugar

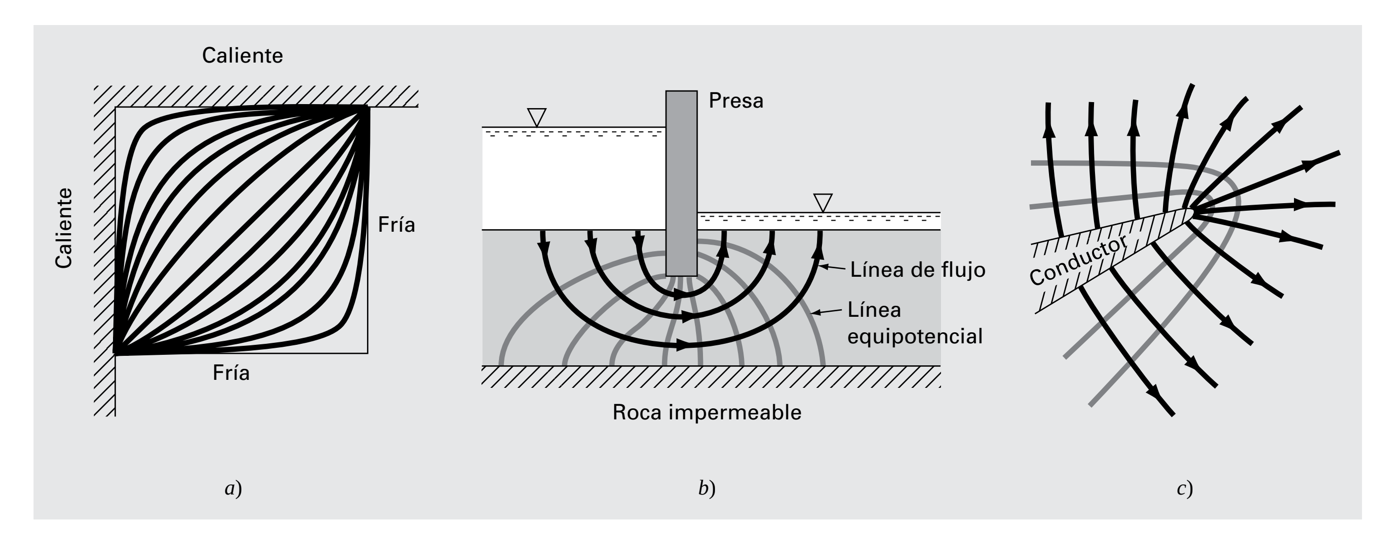


Figura PT8.1
Tres problemas de distribución en estado estacionario que pueden caracterizarse por EDP elípticas. *a*) Distribución de temperatura sobre una placa caliente; *b*) filtración de agua bajo una presa, y *c*) el campo eléctrico cerca del punto de un conductor.

de concentrarse en la distribución en estado estacionario en dos dimensiones espaciales, el problema consiste en determinar cómo la distribución espacial en una dimensión cambia en función del tiempo (figura PT8.2*b*). Así, la solución consiste de una serie de distribuciones espaciales que corresponden al estado de la barra en diferentes momentos. Usando una analogía con la fotografía, la categoría elíptica da una imagen del sistema en estado estacionario; mientras que la categoría parabólica ofrece una película de cómo cambia de un estado a otro. Como con los demás tipos de EDP descritos aquí, las ecuaciones parabólicas son útiles para caracterizar, por analogía, una amplia variedad de otros problemas de ingeniería.

La clase final de EDP, la categoría *hiperbólica*, también tiene que ver con *problemas de propagación*. Sin embargo, una importante diferencia manifestada por la ecuación de onda, en la tabla PT8.1, es que la incógnita se caracteriza por una segunda derivada con respecto al tiempo. En consecuencia, la solución oscila.

La cuerda vibrante de la figura PT8.3 es un modelo físico sencillo que puede describirse por la ecuación de onda. La solución consiste de varios estados característicos en que la cuerda oscila. Varios sistemas de ingeniería (como las vibraciones de barras y vigas, los movimientos de ondas de fluido y la transmisión de señales acústicas y eléctricas) pueden caracterizarse con este modelo.

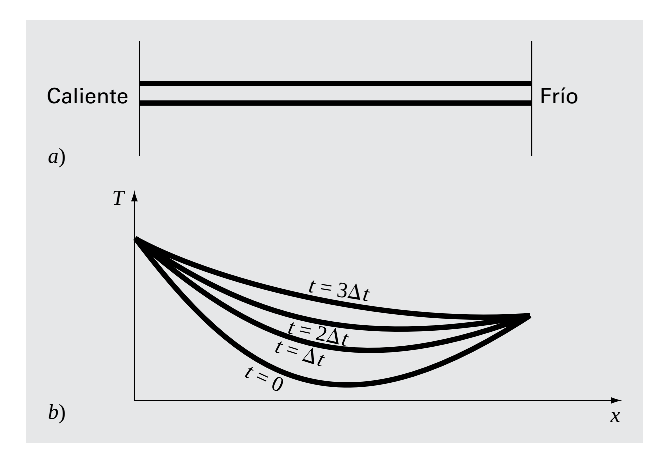


Figura PT8.2
a) Barra larga y delgada que está aislada, excepto en sus extremos. La dinámica de la distribución unidimensional de temperatura a lo largo de la barra puede describirse mediante una EDP parabólica. *b*) La solución, que consiste en distribuciones correspondientes al estado de la barra en diferentes momentos.

PT8.1.2 Métodos anteriores a la computadora para resolver EDP

Antes de la era de las computadoras digitales, los ingenieros dependían de las soluciones analíticas o exactas para las ecuaciones diferenciales parciales. Aparte de los casos más simples, dichas soluciones requerían de un gran esfuerzo y complejidad matemática. Además, muchos sistemas físicos no podían resolverse directamente; tenían que simplificarse utilizando linealizaciones, representaciones geométricas sencillas y otras idealizaciones. Aunque esas soluciones son elegantes y profundas, están limitadas con respecto a la fidelidad para representar sistemas reales (en especial, aquellos que son altamente no lineales y de forma irregular).

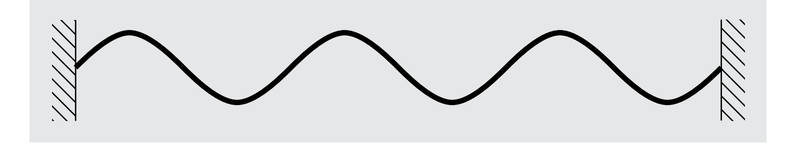


Figura PT8.3
Una cuerda tensa que vibra a baja amplitud es un sistema físico simple que puede caracterizarse por una EDP hiperbólica.

PT8.2 ORIENTACIÓN

Antes de proceder con los métodos numéricos para resolver ecuaciones diferenciales parciales, alguna orientación resultará de utilidad. La siguiente información tiene el propósito de presentarle un panorama general del material analizado en la parte ocho. Además, hemos formulado objetivos para concentrar sus estudios en el tema.

PT8.2.1 Alcance y presentación preliminar

La figura PT8.4 muestra un panorama general de la parte ocho. En esta parte del libro se analizarán dos amplias categorías de métodos numéricos. Los procedimientos por diferencias finitas, que se cubrirán en los capítulos 29 y 30, se basan en la aproximación de la solución en un número finito de puntos. En cambio, los métodos por elementos finitos, que se estudiarán en el capítulo 31, aproximan la solución por partes o "elementos". Varios parámetros se ajustan hasta que las aproximaciones conformen la ecuación diferencial correspondiente en un sentido óptimo.

El *capítulo 29* se dedica a soluciones por *diferencias finitas* de *ecuaciones elípticas*. Antes de poner en práctica los métodos, deducimos la ecuación de Laplace para el problema físico de la distribución de temperatura en una placa caliente. Después, se describe un procedimiento estándar de solución: el *método de Liebmann*. Ilustraremos cómo se utiliza dicho procedimiento para calcular la distribución de la variable escalar principal, la temperatura, así como la de una variable vectorial secundaria: el flujo de calor. La última sección del capítulo se ocupa de las *condiciones de frontera*. Este material comprende procedimientos para diferentes tipos de condiciones, así como para fronteras irregulares.

En el *capítulo 30* se estudian las soluciones por *diferencias finitas* de *ecuaciones parabólicas*. Como en el análisis de ecuaciones elípticas, primero ofreceremos una introducción a un problema físico, la ecuación de conducción del calor en una barra unidimensional. Después presentamos algoritmos implícitos y explícitos para resolver esta ecuación. Luego se analiza un método implícito eficiente y confiable: la *técnica de Crank-Nicolson*. Por último, describimos un procedimiento particularmente efectivo para resolver ecuaciones parabólicas bidimensionales, el *método implícito de dirección alternante*, o *método IDA*.

Observe que hemos omitido las ecuaciones hiperbólicas por estar más allá del alcance de este libro. El epílogo de esta parte del libro contiene referencias relacionadas con este tipo de EDP.

En el *capítulo 31* veremos otro procedimiento fundamental para resolver EDP: el *método del elemento finito*. Como es esencialmente diferente del procedimiento por diferencias finitas, hemos dedicado la sección inicial del capítulo a una visión general. Después mostramos cómo se utiliza el método del elemento finito para calcular la distribución de temperatura en estado estacionario de una barra caliente. Por último, ofrecemos una introducción a algunos de los problemas al extender este análisis a problemas bidimensionales.

El *capítulo 32* se dedica a problemas en todos los campos de la ingeniería. Por último, se presenta una breve sección de repaso al final de la parte ocho. Este epílogo resume información importante relacionada con las EDP. Este material comprende un análisis de las ventajas y las desventajas esenciales para su implementación en la ingeniería. El epílogo también incluye referencias para temas avanzados.

PT8.2.2 Metas y objetivos

Objetivos de estudio Al terminar la parte ocho, deberá haber incrementado su capacidad para enfrentar y resolver ecuaciones diferenciales parciales. Las metas de estudio generales deberán com-

prender el dominio de las técnicas, teniendo la capacidad de evaluar la confiabilidad de las respuestas, y de elegir el "mejor" método (o métodos) para cualquier problema particular. Además de estos objetivos generales, deberán dominarse los objetivos de estudio específicos de la tabla PT8.2.

Objetivos de cómputo Se pueden desarrollar algoritmos computacionales para muchos de los métodos de la parte ocho. Por ejemplo, usted puede encontrar ilustrativo el desarrollo de un programa general, para simular la distribución de la temperatura en estado estacionario sobre una placa caliente. Además, tal vez usted quiera desarrollar programas para implementar el sencillo método explícito y el de Crank-Nicolson, para resolver EDP parabólicas en una dimensión espacial.

Por último, una de sus metas más importantes deberá ser dominar varios de los paquetes de software de uso general ampliamente difundidos. En particular, usted deberá volverse un adepto al uso de esas herramientas para implementar métodos numéricos que resuelvan problemas de ingeniería.

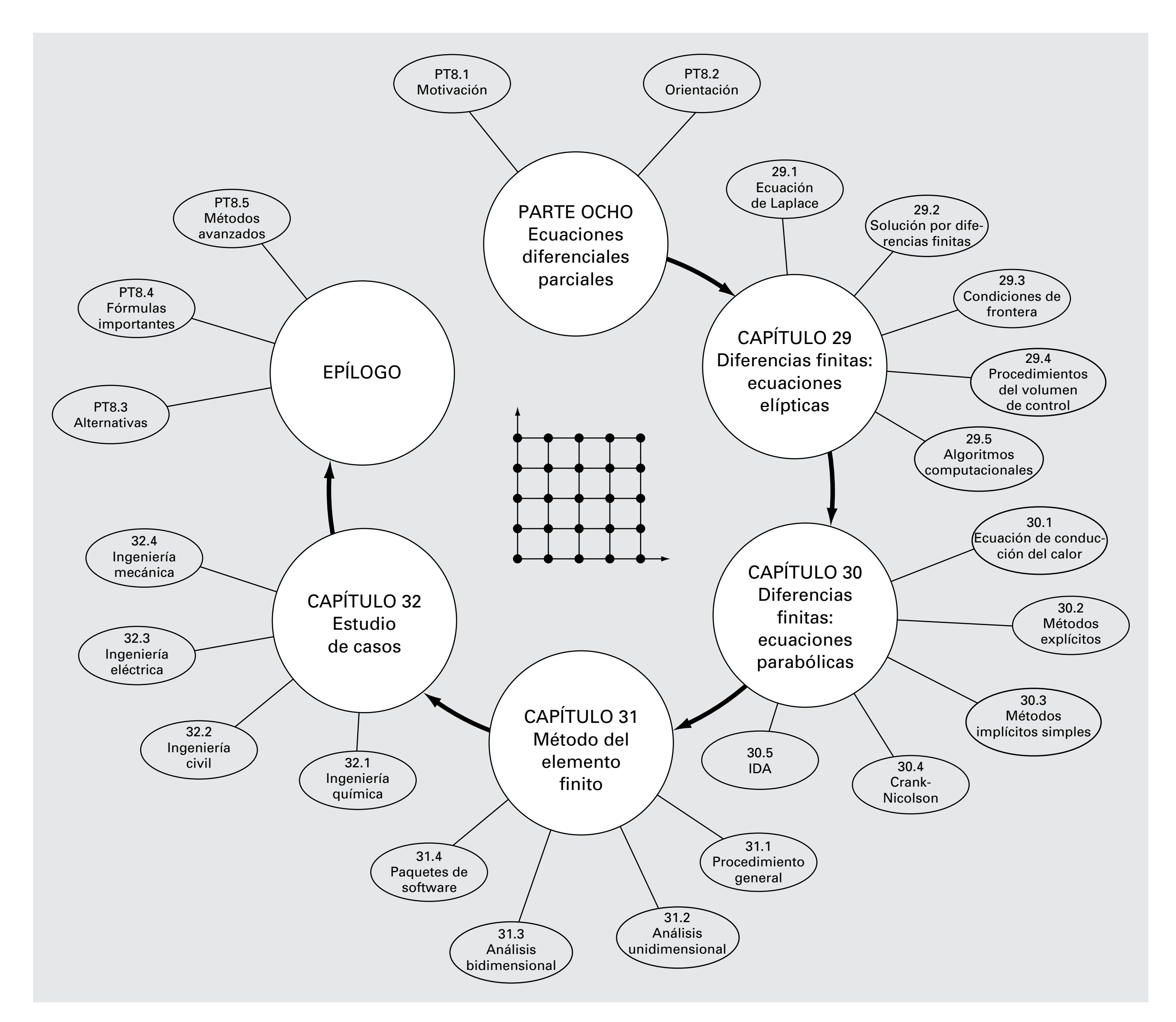


Figura PT8.4
Representación esquemática de la organización del material de la parte ocho: Ecuaciones diferenciales parciales.

Tabla PT8.2 Objetivos específicos de estudio de la parte ocho.

1. Reconocer la diferencia entre las EDP elípticas, parabólicas e hiperbólicas.
2. Comprender la diferencia fundamental entre los procedimientos de diferencias finitas y de elementos finitos.
3. Entender que el método de Liebmann es equivalente al método de Gauss-Seidel para resolver ecuaciones algebraicas lineales simultáneas.
4. Saber cómo determinar variables secundarias para problemas de campos bidimensionales.
5. Distinguir la diferencia entre las condiciones Dirichlet y las condiciones de la derivada en la frontera.
6. Saber cómo usar factores ponderados para incorporar fronteras irregulares en un esquema por diferencias finitas para las EDP.
7. Implementar la aproximación del volumen de control para las soluciones numéricas de las EDP.
8. Conocer la diferencia entre convergencia y estabilidad de EDP parabólicas.
9. Distinguir la diferencia entre esquemas explícitos y esquemas implícitos para resolver EDP parabólicas.
10. Reconocer cómo los criterios de estabilidad para métodos explícitos disminuyen en su utilidad para resolver EDP parabólicas.
11. Saber cómo interpretar moléculas computacionales.
12. Comprender cómo el procedimiento IDA tiene alta eficiencia en la solución de ecuaciones parabólicas en dos dimensiones espaciales.
13. Comprender la diferencia entre el método directo y el método de residuos ponderados para deducir elementos de ecuaciones.
14. Saber cómo implementar el método de Galerkin.
15. Entender los beneficios de la integración por partes durante la deducción de elementos de ecuaciones; en particular, reconocer las implicaciones que se tienen al disminuir la segunda derivada a una primera derivada.

CAPÍTULO 29

Diferencias finitas: ecuaciones elípticas

En ingeniería, las ecuaciones elípticas se usan comúnmente para caracterizar problemas en estado estacionario con valores en la frontera. Antes de mostrar la manera en que se resuelven, ilustraremos cómo se deduce en un caso simple (la ecuación de Laplace), a partir de un problema físico.

29.1 LA ECUACIÓN DE LAPLACE

Como se mencionó en la introducción de esta parte del libro, la ecuación de Laplace se utiliza para modelar diversos problemas que tienen que ver con el potencial de una variable desconocida. Debido a su simplicidad y a su relevancia en la mayoría de las áreas de la ingeniería, usaremos una placa caliente para deducir y resolver esta EDP elíptica. Se emplearán problemas académicos y problemas de la ingeniería (capítulo 32) para ilustrar la aplicabilidad del modelo a otros problemas de ingeniería.

En la figura 29.1 se muestra un elemento sobre la cara de una placa rectangular delgada de espesor Δz. La placa está totalmente aislada excepto en sus extremos, donde la temperatura puede ajustarse a un nivel preestablecido. El aislamiento y el espesor de la placa permiten que la transferencia de calor esté limitada solamente a las dimensiones x y y. En estado estacionario, el flujo de calor hacia el elemento en una unidad de tiempo Δt debe ser igual al flujo de salida, es decir,

$$q(x)\,\Delta y\,\Delta z\,\Delta t + q(y)\,\Delta x\,\Delta z\,\Delta t = q(x+\Delta x)\,\Delta y\,\Delta z\,\Delta t + q(y+\Delta y)\Delta x\,\Delta z\,\Delta t \tag{29.1}$$

donde $q(x)$ y $q(y)$ = los flujos de calor en x y y, respectivamente [cal/(cm^2 · s)]. Dividiendo entre Δz y Δt, y reagrupando términos, se obtiene

$$[q(x) - q(x+\Delta x)]\,\Delta y + [q(y) - q(y+\Delta y)]\Delta x = 0$$

Multiplicando el primer término por $\Delta x/\Delta x$, y el segundo por $\Delta y/\Delta y$ se obtiene

$$\frac{q(x)-q(x+\Delta x)}{\Delta x}\Delta x\;\Delta y + \frac{q(y)-q(y+\Delta y)}{\Delta y}\Delta y\;\Delta x = 0 \tag{29.2}$$

Dividiendo entre $\Delta x\,\Delta y$, y tomando el límite, se llega a

$$-\frac{\partial q}{\partial x} - \frac{\partial q}{\partial y} = 0 \tag{29.3}$$

donde las derivadas parciales resultan de las definiciones en las ecuaciones (PT8.1) y (PT8.2).

La ecuación (29.3) es una ecuación diferencial parcial, que es una expresión de la conservación de la energía en la placa. Sin embargo, la ecuación no puede resolverse, a menos que se especifiquen los flujos de calor en los extremos de la placa. Debido a que se dan condiciones de frontera para la temperatura, la ecuación (29.3) debe reformularse en términos de la temperatura. La relación entre flujo y temperatura está dada por la *ley de Fourier de conducción del calor*, la cual se representa como

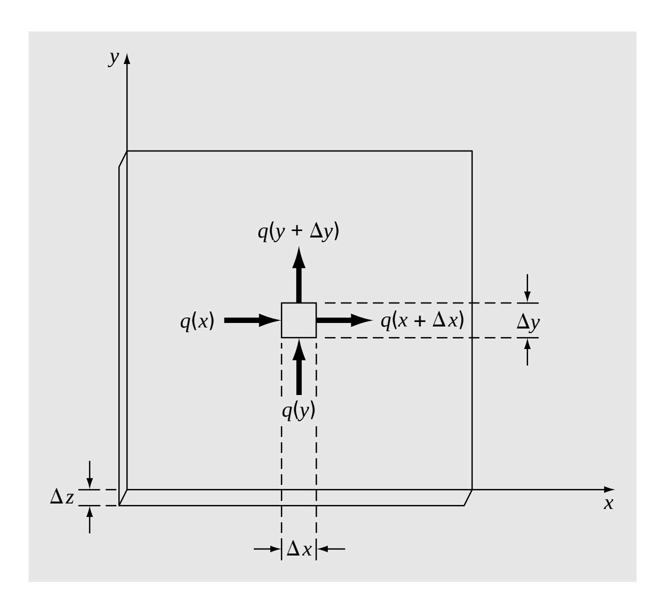


Figura 29.1
Placa delgada de espesor Δz. Se muestra un elemento, con el cual se hace el balance de calor.

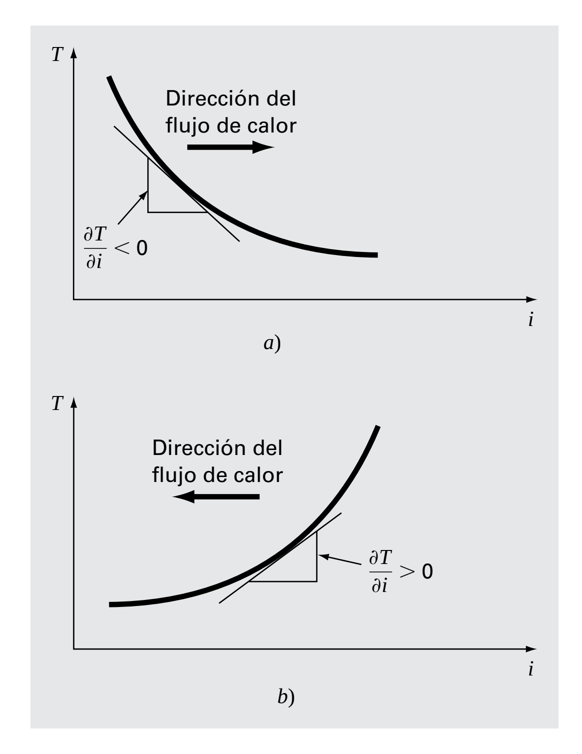


Figura 29.2
Representación gráfica de un gradiente de temperatura. Debido a que el calor se transfiere hacia abajo desde una temperatura alta a una baja, el flujo en *a*) va de izquierda a derecha en la dirección *i* positiva. Sin embargo, debido a la orientación de las coordenadas cartesianas, la pendiente es negativa en este caso. Es decir, un gradiente negativo se relaciona con un flujo positivo. Éste es el origen del signo menos en la ley de Fourier de conducción de calor. El caso inverso se ilustra en *b*), donde el gradiente positivo se relaciona con un flujo de calor negativo de derecha a izquierda.

$$q_i = -k\rho C\frac{\partial T}{\partial i} \tag{29.4}$$

donde q_i = flujo de calor en la dirección de la dimensión i [cal/(cm² · s)], k = coeficiente de *difusividad térmica* (cm²/s), ρ = densidad del material (g/cm³), C = capacidad calorífica del material [cal/(g · °C)] y T = temperatura (°C), que se define como

$$T = \frac{H}{\rho CV}$$

donde H = calor (cal) y V = volumen (cm³). Algunas veces, el término que está multiplicando a la derivada parcial en la ecuación (29.4) se trata como un solo término,

$$k' = k\rho C \tag{29.5}$$

donde k' se conoce como el *coeficiente de conductividad térmica* [cal/(s · cm · °C)]. En ambos casos, k y k' son parámetros que determinan qué tan bien conduce calor el material.

A la ley de Fourier algunas veces se le llama *ecuación constitutiva*. Esta connotación se le da porque proporciona un mecanismo que define las interacciones internas del sistema. Una inspección de la ecuación (29.4) indica que la ley de Fourier especifica que el flujo de calor perpendicular al eje i es proporcional al gradiente o pendiente de la temperatura en la dirección i. El signo negativo asegura que un flujo positivo en la dirección i resulta de una pendiente negativa de alta a baja temperatura (figura 29.2). Sustituyendo la ecuación (29.4) en la ecuación (29.3), se obtiene

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = 0 \tag{29.6}$$

que es la *ecuación de Laplace*. Observe que en el caso donde hay fuentes o pérdidas de calor dentro del dominio bidimensional, la ecuación se puede representar como

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = f(x, y) \tag{29.7}$$

donde $f(x, y)$ es una función que describe las fuentes o pérdidas de calor. La ecuación (29.7) se conoce como *ecuación de Poisson*.

29.2 TÉCNICA DE SOLUCIÓN

Para la solución numérica de las EDP elípticas, como la ecuación de Laplace, se procede en dirección contraria a como se dedujo la ecuación (29.6) en la sección anterior. Recuerde que la deducción de la ecuación (29.6) emplea un balance alrededor de un elemento discreto para obtener una ecuación algebraica en diferencias, que caracteriza el flujo de calor para una placa. Tomando el límite, esta ecuación en diferencias se convirtió en una ecuación diferencial [ecuación (29.3)].

En la solución numérica, las representaciones por diferencias finitas basadas en tratar la placa como una malla de puntos discretos (figura 29.3) se sustituyen por las derivadas parciales en la ecuación (29.6). Como se describe a continuación, la EDP se transforma en una ecuación algebraica en diferencias.

29.2.1 La ecuación laplaciana en diferencias

Las diferencias centrales basadas en el esquema de malla de la figura 29.3 son (véase figura 23.3)

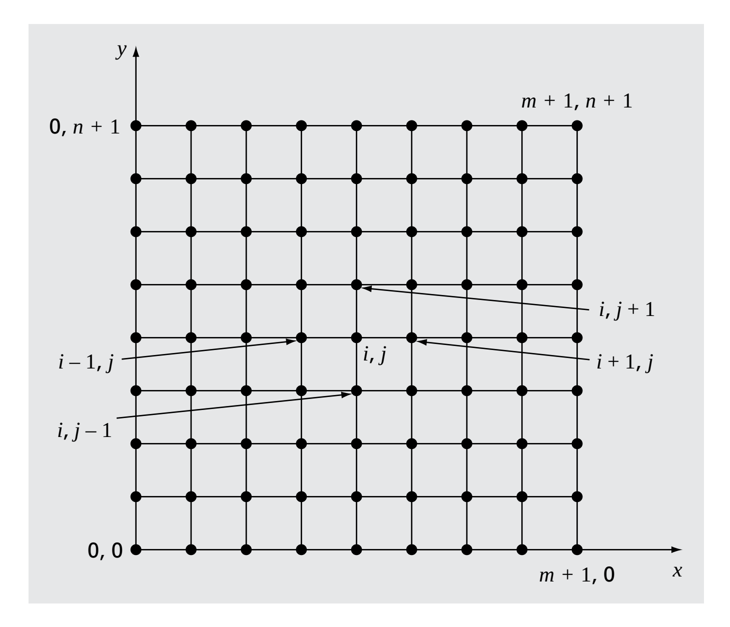


Figura 29.3
Malla usada para la solución por diferencias finitas de las EDP elípticas en dos variables independientes, como la ecuación de Laplace.

$$\frac{\partial^2 T}{\partial x^2} = \frac{T_{i+1,j} - 2T_{i,j} + T_{i-1,j}}{\Delta x^2}$$

y

$$\frac{\partial^2 T}{\partial y^2} = \frac{T_{i,j+1} - 2T_{i,j} + T_{i,j-1}}{\Delta y^2}$$

las cuales tienen errores de $O[\Delta(x)^2]$ y $O[\Delta(y)^2]$, respectivamente. Sustituyendo estas expresiones en la ecuación (29.6) se obtiene

$$\frac{T_{i+1,j} - 2T_{i,j} + T_{i-1,j}}{\Delta x^2} + \frac{T_{i,j+1} - 2T_{i,j} + T_{i,j-1}}{\Delta y^2} = 0$$

En la malla cuadrada de la figura 29.3, $\Delta x = \Delta y$, y reagrupando términos, la ecuación se convierte en

$$T_{i+1,j} + T_{i-1,j} + T_{i,j+1} + T_{i,j-1} - 4T_{i,j} = 0 \tag{29.8}$$

Esta relación, que se satisface por todos los puntos interiores de la placa, se conoce como *ecuación laplaciana en diferencias*.

Además, se deben especificar las condiciones de frontera en los extremos de la placa para obtener una solución única. El caso más simple es aquel donde la temperatura en la frontera es un valor fijo. Ésta se conoce como *condición de frontera de Dirichlet*. Tal es el caso de la figura 29.4, donde los extremos se mantienen a temperaturas constantes. En el caso ilustrado en la figura 29.4, un balance en el nodo (1, 1) es, de acuerdo con la ecuación (29.8),

$$T_{21} + T_{01} + T_{12} + T_{10} - 4T_{11} = 0 \tag{29.9}$$

Sin embargo, $T_{01} = 75$ y $T_{10} = 0$, y, por lo tanto, la ecuación (29.9) se expresa como

$$-4T_{11} + T_{12} + T_{21} = -75$$

Ecuaciones similares se pueden desarrollar para los otros puntos interiores. El resultado es el siguiente conjunto de nueve ecuaciones simultáneas con nueve incógnitas:

$$\begin{array}{ccccccccccc}
4T_{11} & -T_{21} & & -T_{12} & & & & & & = & 75 \\
-T_{11} & +4T_{21} & -T_{31} & & -T_{22} & & & & & = & 0 \\
 & -T_{21} & +4T_{31} & & & -T_{32} & & & & = & 50 \\
-T_{11} & & & +4T_{12} & -T_{22} & & -T_{13} & & & = & 75 \\
 & -T_{21} & & -T_{12} & +4T_{22} & -T_{32} & & -T_{23} & & = & 0 \\
 & & -T_{31} & & -T_{22} & +4T_{32} & & & -T_{33} & = & 50 \\
 & & & -T_{12} & & & +4T_{13} & -T_{23} & & = & 175 \\
 & & & & -T_{22} & & -T_{13} & +4T_{23} & -T_{33} & = & 100 \\
 & & & & & -T_{32} & & -T_{23} & +4T_{33} & = & 150
\end{array} \tag{29.10}$$

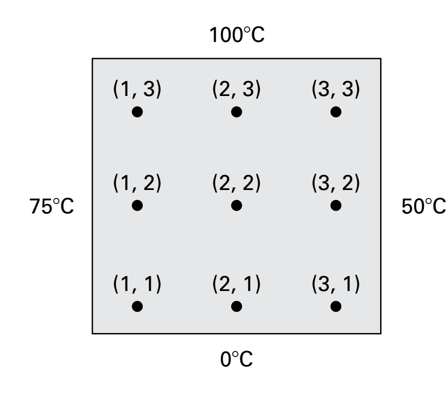


Figura 29.4
Una placa caliente donde las temperaturas en la frontera se mantienen a niveles constantes. Este caso se denomina condición de frontera de Dirichlet.

29.2.2 El método de Liebmann

En la mayoría de las soluciones numéricas de la ecuación de Laplace se tienen sistemas que son mucho más grandes que la ecuación (29.10). Por ejemplo, para una malla de 10 por 10 se tienen 100 ecuaciones algebraicas lineales. En la parte tres se analizaron técnicas de solución para estos tipos de ecuaciones.

Observe que hay un máximo de cinco incógnitas por línea en la ecuación (29.10). Para mallas grandes se encuentra que un número significativo de los términos será igual a cero. Cuando se aplican los métodos de eliminación con toda la matriz a estos sistemas dispersos, se ocupa una gran cantidad de memoria de la computadora, almacenando ceros. Por esta razón, los métodos aproximados representan un mejor procedimiento para obtener soluciones de EDP elípticas. El método comúnmente empleado es el de *Gauss-Seidel*, el cual, cuando se aplica a las EDP, también se conoce como el *método de Liebmann*. Con esta técnica, la ecuación (29.8) se expresa como

$$T_{i,j} = \frac{T_{i+1,j} + T_{i-1,j} + T_{i,j+1} + T_{i,j-1}}{4} \tag{29.11}$$

y se resuelve de manera iterativa para $j = 1$ hasta n e $i = 1$ hasta m. Como la ecuación (29.8) es diagonalmente dominante, este procedimiento al final convergerá a una solución estable (recuerde la sección 11.2.1). Algunas veces se utiliza la sobrerrelajación para acelerar la velocidad de la convergencia, aplicando la siguiente fórmula después de cada iteración:

$$T_{i,j}^{\text{nuevo}} = \lambda T_{i,j}^{\text{nuevo}} + (1-\lambda) T_{i,j}^{\text{anterior}} \tag{29.12}$$

donde $T_{i,j}^{\text{nuevo}}$ y $T_{i,j}^{\text{anterior}}$ son los valores de $T_{i,j}$ de la actual iteración y de la previa, respectivamente; λ es un factor de ponderación que está entre 1 y 2.

Como en el método convencional de Gauss-Seidel, las iteraciones se repiten hasta que los valores absolutos de todos los errores relativos porcentuales $(\varepsilon_a)_{i,j}$ están por debajo de un criterio preespecificado de terminación ε_s. Dichos errores relativos porcentuales se estiman mediante

$$\left|(\varepsilon_a)_{i,j}\right| = \left|\frac{T_{i,j}^{\text{nuevo}} - T_{i,j}^{\text{anterior}}}{T_{i,j}^{\text{nuevo}}}\right| 100\% \tag{29.13}$$

EJEMPLO 29.1 Temperatura de una placa caliente con condiciones de frontera fijas

Planteamiento del problema Con el método de Liebmann (Gauss-Seidel) calcule la temperatura de la placa caliente de la figura 29.4. Emplee la sobrerrelajación con un valor de 1.5 para el factor de ponderación, e itere hasta $\varepsilon_s = 1\%$.

Solución La ecuación (29.11) en $i = 1$, $j = 1$ es

$$T_{11} = \frac{0 + 75 + 0 + 0}{4} = 18.75$$

y aplicando sobrerrelajación se obtiene

$$T_{11} = 1.5(18.75) + (1 - 1.5)0 = 28.125$$

Para $i = 2$, $j = 1$,

$$T_{21} = \frac{0 + 28.125 + 0 + 0}{4} = 7.03125$$

$$T_{21} = 1.5(7.03125) + (1 - 1.5)0 = 10.54688$$

Para $i = 3$, $j = 1$,

$$T_{31} = \frac{50 + 10.54688 + 0 + 0}{4} = 15.13672$$

$$T_{31} = 1.5(15.13672) + (1 - 1.5)0 = 22.70508$$

100°C

75°C				50°C
	78.59	76.06	69.71	
75°C	63.21	56.11	52.34	50°C
	43.00	33.30	33.89	

0°C

Figura 29.5
Distribución de temperatura en una placa caliente, sujeta a condiciones de frontera fijas.

El cálculo se repite con los otros renglones:

$T_{12} = 38.67188$ $T_{22} = 18.45703$ $T_{32} = 34.18579$

$T_{13} = 80.12696$ $T_{23} = 74.46900$ $T_{33} = 96.99554$

Como todos los $T_{i,j}$ son inicialmente cero, entonces todos los ε_a para la primera iteración serán 100%.

En la segunda iteración, los resultados son:

$T_{11} = 32.51953$ $T_{21} = 22.35718$ $T_{31} = 28.60108$

$T_{12} = 57.95288$ $T_{22} = 61.63333$ $T_{32} = 71.86833$

$T_{13} = 75.21973$ $T_{23} = 87.95872$ $T_{33} = 67.68736$

El error para $T_{1,1}$ se estima como sigue [ecuación (29.13)]

$$|(\varepsilon_a)_{1,1}| = \left|\frac{32.51953 - 28.12500}{32.51953}\right| 100\% = 13.5\%$$

Debido a que este valor está por arriba del criterio de terminación de 1%, se continúa el cálculo. La novena iteración da como resultado

$T_{11} = 43.00061$ $T_{21} = 33.29755$ $T_{31} = 33.88506$

$T_{12} = 63.21152$ $T_{22} = 56.11238$ $T_{32} = 52.33999$

$T_{13} = 78.58718$ $T_{23} = 76.06402$ $T_{33} = 69.71050$

donde el error máximo es 0.71%.

En la figura 29.5 se muestran los resultados. Como se esperaba, se ha establecido un gradiente al fluir el calor de altas a bajas temperaturas.

29.2.3 Variables secundarias

Como la distribución de temperatura está descrita por la ecuación de Laplace, ésta se considera la variable principal en el problema de la placa caliente. En este caso, así como en otros problemas donde se tengan EDP, las variables secundarias también pueden ser importantes. De hecho, en ciertos contextos de ingeniería, la variable secundaria puede realmente ser más importante.

En la placa caliente, una variable secundaria es el flujo de calor a través de la superficie de la placa. Esta cantidad se calcula a partir de la ley de Fourier. Las aproximaciones por diferencias finitas centradas para las primeras derivadas (recuerde la figura 23.3) se sustituyen en la ecuación (29.4) para obtener los siguientes valores del flujo de calor en las dimensiones x y y:

$$q_x = -k' \frac{T_{i+1,j} - T_{i-1,j}}{2\,\Delta x} \tag{29.14}$$

y

$$q_x = -k' \frac{T_{i,j+1} - T_{i,j-1}}{2\,\Delta y} \tag{29.15}$$

El flujo de calor resultante se calcula a partir de estas dos cantidades mediante

$$q_n = \sqrt{q_x^2 + q_y^2} \tag{29.16}$$

donde la dirección de q_n está dada por

$$\theta = \tan^{-1}\left(\frac{q_y}{q_x}\right) \tag{29.17}$$

para $q_x > 0$ y

$$\theta = \tan^{-1}\left(\frac{q_y}{q_x}\right) + \pi \tag{29.18}$$

para $q_x < 0$. Recuerde que el ángulo puede expresarse en grados multiplicándolo por $180°/\pi$. Si $q_x = 0$, θ es $\pi/2$ (90°) o $3\pi/2$ (270°), según q_y sea positivo o negativo, respectivamente.

EJEMPLO 29.2 Distribución de flujo en una placa caliente

Planteamiento del problema Empleando los resultados del ejemplo 29.1 determine la distribución del flujo de calor en la placa caliente de la figura 29.4. Suponga que la placa es de 40 × 40 cm y que está hecha de aluminio [$k' = 0.49$ cal/(s · cm · °C)].

Solución Para $i = j = 1$, la ecuación (29.14) se utiliza para calcular

$$q_x = -0.49 \frac{\text{cal}}{\text{s} \cdot \text{cm} \cdot °\text{C}} \frac{(33.29755 - 75)°\text{C}}{2(10\ \text{cm})} = 1.022\ \text{cal/(cm}^2 \cdot \text{s)}$$

y [de la ecuación (29.15)]

$$q_y = -0.49 \frac{\text{cal}}{\text{s} \cdot \text{cm} \cdot °\text{C}} \frac{(63.21152 - 0)°\text{C}}{2(10\ \text{cm})} = -1.549\ \text{cal/(cm}^2 \cdot \text{s)}$$

El flujo resultante se calcula con la ecuación (29.16):

$$q_n = \sqrt{(1.022)^2 + (-1.549)^2} = 1.856\ \text{cal/(cm}^2 \cdot \text{s)}$$

y el ángulo de su trayectoria mediante la ecuación (29.17)

$$\theta = \tan^{-1}\left(\frac{-1.549}{1.022}\right) = -0.98758 \times \frac{180°}{\pi} = -56.584°$$

Así, en este punto, el flujo de calor está dirigido hacia abajo y a la derecha. Pueden calcularse los valores en otros puntos de la malla; los resultados se muestran en la figura 29.6.

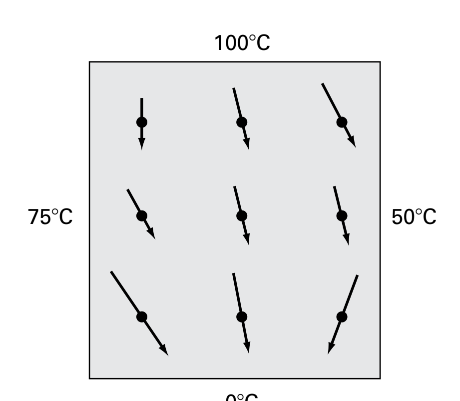


Figura 29.6
Flujo de calor en una placa sujeta a temperaturas fijas en las fronteras. Observe que la longitud de las flechas es proporcional a la magnitud del flujo.

29.3 CONDICIONES EN LA FRONTERA

Debido a que está libre de complicaciones, la placa rectangular con condiciones de frontera fijas representa un ideal para mostrar cómo se resuelven numéricamente las EDP elípticas. Ahora veremos otro problema que ampliará nuestras habilidades para abordar problemas más realistas. Éste considera fronteras en donde se especifica la derivada, y fronteras que tienen forma irregular.

29.3.1 Condiciones con derivada en la frontera

La condición de frontera fija o de Dirichlet analizada hasta ahora es uno de los diferentes tipos usados en las ecuaciones diferenciales parciales. Una alternativa común es el caso donde se da la derivada, que se conoce comúnmente como una *condición de frontera de Neumann*. En el problema de la placa caliente, esto corresponde a especificar el flujo de calor, más que la temperatura en la frontera. Un ejemplo es la situación donde el extremo está aislado. En tal caso, la derivada es cero. Esta conclusión se obtiene directamente de la ecuación (29.4), ya que aislar una frontera significa que el flujo de calor (y, en consecuencia, el gradiente) debe ser cero. Otro ejemplo sería el caso donde se pierde calor a través del extremo por mecanismos predecibles, como radiación y convección.

$T_{0,j+1}$
$T_{-1,j}$ $T_{0,j}$ $T_{1,j}$
$T_{0,j-1}$

Figura 29.7
Un nodo frontera (0, *j*) en el extremo izquierdo de una placa caliente. Para aproximar la derivada normal al extremo (es decir, la derivada *x*), se localiza un punto imaginario (–1, *j*) a una distancia Δx más allá del extremo.

En la figura 29.7 se muestra un nodo (0, *j*) en el extremo izquierdo de una placa caliente. Aplicando la ecuación (29.8) en este punto, se obtiene

$$T_{1,j} + T_{-1,j} + T_{0,j+1} + T_{0,j-1} - 4T_{0,j} = 0 \tag{29.19}$$

Observe que para esta ecuación se necesita un punto imaginario (–1, *j*) que esté fuera de la placa. Aunque este punto exterior ficticio podría parecer que representa un problema, realmente sirve para incorporar la derivada de la condición de frontera en el problema, lo cual se logra representando la primera derivada en la dimensión *x* en (0, *j*) por la diferencia dividida finita

$$\frac{\partial T}{\partial x} \cong \frac{T_{1,j} - T_{-1,j}}{2\,\Delta x}$$

donde se puede despejar

$$T_{-1,j} = T_{1,j} - 2\,\Delta x \frac{\partial T}{\partial x}$$

Ahora se tiene una relación para $T_{-1,j}$ que incluye la derivada. Esta relación se sustituye en la ecuación (29.19) para obtener

$$2T_{1,j} - 2\,\Delta x \frac{\partial T}{\partial x} + T_{0,j+1} + T_{0,j-1} - 4T_{0,j} = 0 \tag{29.20}$$

Así, hemos incorporado la derivada en la ecuación.

Es posible desarrollar relaciones similares para las condiciones de frontera con derivadas en los otros extremos. El siguiente ejemplo muestra cómo llevarlo a cabo en la placa caliente.

EJEMPLO 29.3 Placa caliente con un extremo aislado

Planteamiento del problema Repita el mismo problema del ejemplo 29.1, pero con el extremo inferior aislado.

Solución La ecuación general que caracteriza una derivada en el extremo inferior (es decir, en *j* = 0) en una placa caliente es

$$T_{i+1,0} + T_{i-1,0} + 2T_{i,1} - 2\,\Delta y \frac{\partial T}{\partial y} - 4T_{i,0} = 0$$

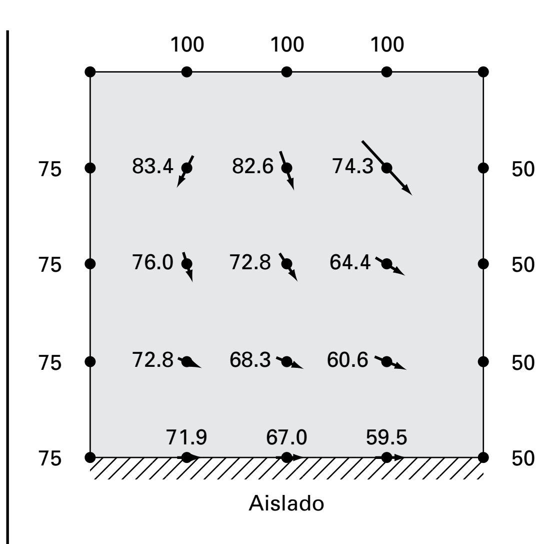


Figura 29.8
Temperatura y distribución de flujo en una placa caliente sujeta a condiciones de frontera fijas, excepto en un extremo inferior aislado.

En el extremo aislado, la derivada es cero y la ecuación se convierte en

$$T_{i+1,0} + T_{i-1,0} + 2T_{i,1} - 4T_{i,0} = 0$$

Las ecuaciones simultáneas para la distribución de temperatura en la placa de la figura 29.4 con un extremo inferior aislado se escribe en forma matricial como

$$\begin{bmatrix} 4 & -1 & & -2 & & & & & & & & \\ -1 & 4 & -1 & & -2 & & & & & & & \\ & -1 & 4 & & & -2 & & & & & & \\ -1 & & & 4 & -1 & & -1 & & & & & \\ & -1 & & -1 & 4 & -1 & & -1 & & & & \\ & & -1 & & -1 & 4 & & & -1 & & & \\ & & & -1 & & & 4 & -1 & & -1 & & \\ & & & & -1 & & -1 & 4 & -1 & & -1 & \\ & & & & & -1 & & -1 & 4 & & & -1 \\ & & & & & & -1 & & & 4 & -1 & \\ & & & & & & & -1 & & -1 & 4 & -1 \\ & & & & & & & & -1 & & -1 & 4 \end{bmatrix} \begin{Bmatrix} T_{10} \\ T_{20} \\ T_{30} \\ T_{11} \\ T_{21} \\ T_{31} \\ T_{12} \\ T_{22} \\ T_{32} \\ T_{13} \\ T_{23} \\ T_{33} \end{Bmatrix} = \begin{Bmatrix} 75 \\ 0 \\ 50 \\ 75 \\ 0 \\ 50 \\ 75 \\ 0 \\ 50 \\ 175 \\ 100 \\ 150 \end{Bmatrix}$$

Observe que, debido a las derivadas en las condiciones de frontera, la matriz aumentó de tamaño a 12×12, a diferencia del sistema de 9×9 de la ecuación (29.10), para considerar las tres temperaturas desconocidas del extremo inferior de la placa. De estas ecuaciones se obtiene

$$\begin{array}{lll} T_{10} = 71.91 & T_{20} = 67.01 & T_{30} = 59.54 \\ T_{11} = 72.81 & T_{21} = 68.31 & T_{31} = 60.57 \\ T_{12} = 76.01 & T_{22} = 72.84 & T_{32} = 64.42 \\ T_{13} = 83.41 & T_{23} = 82.63 & T_{33} = 74.26 \end{array}$$

Esos resultados y los flujos calculados (con los mismos parámetros que en el ejemplo 29.2) se muestran en la figura 29.8. Observe que, debido a que el extremo inferior está aislado, la temperatura de la placa es más alta que en la figura 29.5, donde la temperatura del extremo inferior se fijó en cero. Además, el flujo de calor (a diferencia de la figura 29.6) ahora está desviado a la derecha y se mueve paralelamente a la pared aislada.

29.3.2 Fronteras irregulares

Aunque la placa rectangular de la figura 29.4 nos sirve para ilustrar los aspectos fundamentales en la solución de las EDP elípticas, muchos problemas de ingeniería no muestran esa geometría idealizada. Por ejemplo, muchos sistemas tienen fronteras irregulares (figura 29.9).

La figura 29.9 es un sistema útil para ilustrar cómo se pueden tratar las fronteras no rectangulares. Como se muestra, la frontera inferior izquierda de la placa es circular. Observe que tenemos parámetros adicionales ($\alpha_1, \alpha_2, \beta_1, \beta_2$) en cada una de las longitudes que rodean al nodo. Por supuesto que, para la placa mostrada en la figura 29.9, $\alpha_2 = \beta_2 = 1$. Conservaremos estos parámetros en la siguiente deducción, de tal modo que la ecuación resultante sea aplicable a cualquier frontera irregular (y no sólo a la esquina inferior izquierda de una placa caliente). Las primeras derivadas en la dimensión x se aproximan como sigue

$$\left(\frac{\partial T}{\partial x}\right)_{i-1,i} \cong \frac{T_{i,j} - T_{i-1,j}}{\alpha_1 \,\Delta x} \tag{29.21}$$

y

$$\left(\frac{\partial T}{\partial x}\right)_{i,i+1} \cong \frac{T_{i+1,j} - T_{i,j}}{\alpha_2 \,\Delta x} \tag{29.22}$$

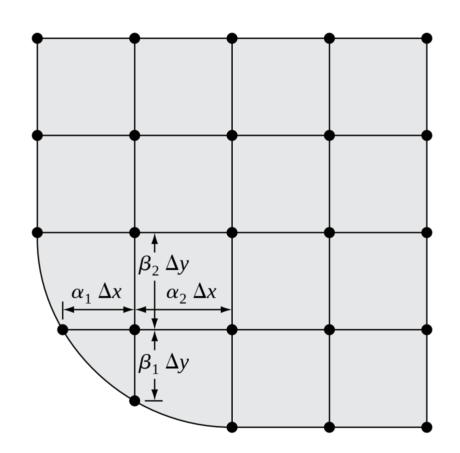


Figura 29.9
Malla de una placa caliente con una frontera en forma irregular. Observe cómo se utilizan los coeficientes ponderados al considerar el espaciamiento no uniforme en la cercanía de la frontera no rectangular.

Las segundas derivadas se obtienen a partir de estas primeras derivadas. Para la dimensión x, la segunda derivada es

$$\frac{\partial^2 T}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial T}{\partial x}\right) = \frac{\left(\frac{\partial T}{\partial x}\right)_{i,i+1} - \left(\frac{\partial T}{\partial x}\right)_{i-1,i}}{\dfrac{\alpha_1\,\Delta x + \alpha_2\,\Delta x}{2}} \tag{29.23}$$

Sustituyendo las ecuaciones (29.21) y (29.22) en la (29.23), obtenemos

$$\frac{\partial^2 T}{\partial x^2} = 2\frac{\dfrac{T_{i-1,j} - T_{i,j}}{\alpha_1\,\Delta x} - \dfrac{T_{i+1,j} - T_{i,j}}{\alpha_2\,\Delta x}}{\alpha_1\,\Delta x + \alpha_2\,\Delta x}$$

Agrupando términos,

$$\frac{\partial^2 T}{\partial x^2} = \frac{2}{\Delta x^2}\left[\frac{T_{i-1,j} - T_{i,j}}{\alpha_1(\alpha_1 + \alpha_2)} + \frac{T_{i+1,j} - T_{i,j}}{\alpha_2(\alpha_1 + \alpha_2)}\right]$$

Es posible desarrollar una ecuación similar en la dimensión y:

$$\frac{\partial^2 T}{\partial y^2} = \frac{2}{\Delta y^2}\left[\frac{T_{i,j-1} - T_{i,j}}{\beta_1(\beta_1 + \beta_2)} + \frac{T_{i,j+1} - T_{i,j}}{\beta_2(\beta_1 + \beta_2)}\right]$$

Sustituyendo estas ecuaciones en la ecuación (29.6), obtenemos

$$\frac{2}{\Delta x^2}\left[\frac{T_{i-1,j} - T_{i,j}}{\alpha_1(\alpha_1 + \alpha_2)} + \frac{T_{i+1,j} - T_{i,j}}{\alpha_2(\alpha_1 + \alpha_2)}\right] + \frac{2}{\Delta y^2}\left[\frac{T_{i,j-1} - T_{i,j}}{\beta_1(\beta_1 + \beta_2)} + \frac{T_{i,j+1} - T_{i,j}}{\beta_2(\beta_1 + \beta_2)}\right] = 0 \tag{29.24}$$

Como se ilustra en el siguiente ejemplo, la ecuación (29.24) se aplica a cualquier nodo que sea adyacente a una frontera irregular de tipo Dirichlet.

EJEMPLO 29.4 Placa caliente con una frontera irregular

Planteamiento del problema Repita el mismo problema del ejemplo 29.1, pero ahora el extremo inferior tendrá la forma que se ilustra en la figura 29.9.

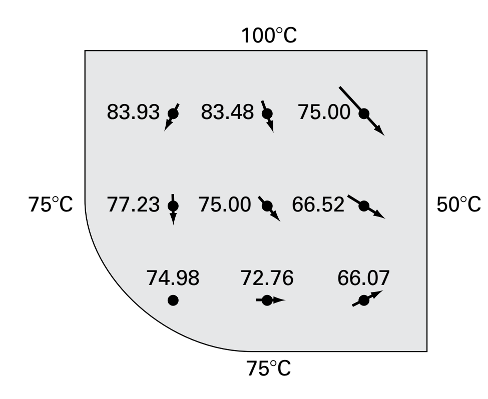


Figura 29.10
Distribución de temperatura y flujo en una placa caliente con una frontera circular.

Solución En el caso de la figura 29.9, $\Delta x = \Delta y$, $\alpha_1 = \beta_1 = 0.732$ y $\alpha_2 = \beta_2 = 1$. Sustituyendo estos valores en la ecuación (29.24), se obtiene la siguiente ecuación para el nodo (1, 1):

$$0.788675(T_{01} - T_{11}) + 0.57735(T_{21} - T_{11}) + 0.788675(T_{10} - T_{11}) + 0.57735(T_{12} - T_{11}) = 0$$

Agrupando términos, esta ecuación se expresa como

$$-4T_{11} + 0.8453T_{21} + 0.8453T_{12} = -1.1547T_{01} - 1.1547T_{10}$$

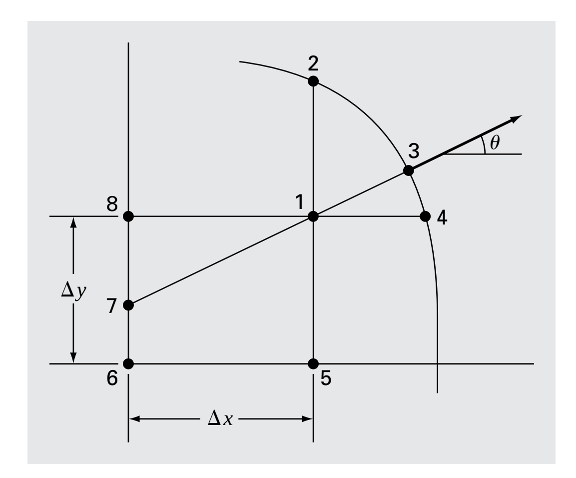
Las ecuaciones simultáneas de la distribución de temperatura sobre la placa de la figura 29.9 con una temperatura en la frontera inferior de 75, se escriben en forma matricial como

$$\begin{bmatrix} 4 & -0.845 & & -0.845 & & & & & \\ -1 & 4 & -1 & & -1 & & & & \\ & -1 & 4 & & & -1 & & & \\ -1 & & & 4 & -1 & & -1 & & \\ & -1 & & -1 & 4 & -1 & & -1 & \\ & & -1 & & -1 & 4 & & & -1 \\ & & & -1 & & & 4 & -1 & \\ & & & & -1 & & -1 & 4 & -1 \\ & & & & & -1 & & -1 & -4 \end{bmatrix} \begin{Bmatrix} T_{11} \\ T_{21} \\ T_{31} \\ T_{12} \\ T_{22} \\ T_{32} \\ T_{13} \\ T_{23} \\ T_{33} \end{Bmatrix} = \begin{Bmatrix} 173.2 \\ 75 \\ 125 \\ 75 \\ 0 \\ 50 \\ 175 \\ 100 \\ 150 \end{Bmatrix}$$

De estas ecuaciones se llega a

$$T_{11} = 74.98 \qquad T_{21} = 72.76 \qquad T_{31} = 66.07$$
$$T_{12} = 74.23 \qquad T_{22} = 75.00 \qquad T_{32} = 66.52$$
$$T_{13} = 83.93 \qquad T_{23} = 83.48 \qquad T_{33} = 75.00$$

Estos resultados, junto con los flujos calculados, se muestran en la figura 29.10. Observe que los flujos se calculan de la misma manera que en la sección 29.2.3, excepto que $(\alpha_1 + \alpha_2)$ y $(\beta_1 + \beta_2)$ se sustituyen por los 2 en los denominadores de las ecuaciones (29.14) y (29.15), respectivamente. En la sección 32.3 se ilustra cómo llevarlo a cabo.



Las derivadas en las condiciones de frontera de forma irregular son más difíciles de formular. En la figura 29.11 se muestra un punto cercano a una frontera irregular donde se especifica la derivada normal.

La derivada normal en el nodo 3 se aproxima por el gradiente entre los nodos 1 y 7,

$$\left.\frac{\partial T}{\partial \eta}\right|_3 = \frac{T_1 - T_7}{L_{17}} \qquad (29.25)$$

Cuando θ es menor a 45°, como se muestra, la distancia del nodo 7 al 8 es $\Delta x \tan \theta$, y se utiliza la interpolación lineal para estimar

Figura 29.11
Frontera curvada donde se especifica el gradiente normal.

$$T_7 = T_8 + (T_6 - T_8)\frac{\Delta x \tan\theta}{\Delta y}$$

La longitud L_{17} es igual a $\Delta x/\cos\theta$. Esta longitud, junto con la aproximación para T_7, puede sustituirse en la ecuación (29.25) para obtener

$$T_1 = \left(\frac{\Delta x}{\cos\theta}\right)\frac{\partial T}{\partial \eta}\bigg|_3 + T_6\frac{\Delta x \tan\theta}{\Delta y} + T_8\left(1 - \frac{\Delta x \tan\theta}{\Delta y}\right) \tag{29.26}$$

Tal ecuación proporciona un medio para incorporar el gradiente normal en el método de diferencias finitas. En los casos donde θ es mayor a 45°, deberá usarse una ecuación diferente. La determinación de esta fórmula se deja como ejercicio para el lector.

29.4 EL MÉTODO DEL VOLUMEN DE CONTROL

Para resumir, el método por diferencias finitas o series de Taylor divide al continuo en nodos (figura 29.12*a*). La ecuación diferencial parcial correspondiente se escribe para cada uno de estos nodos. Las aproximaciones por diferencias finitas, entonces, se sustituyen por las derivadas para llevar las ecuaciones a una forma algebraica.

Un procedimiento de esto es bastante simple y directo con mallas ortogonales (es decir, rectangulares) y coeficientes constantes. Sin embargo, este procedimiento se vuelve un poco más complicado cuando se tiene derivadas como condiciones de fronteras con forma irregular.

En la figura 29.13 se muestra un ejemplo de un sistema donde se presentan dificultades adicionales. La placa está hecha de dos materiales diferentes y los espacios en la malla son distintos. Además, la mitad de su extremo superior está sujeta a transferencia de calor por convección; mientras que la otra mitad está aislada. Obtener las ecuaciones para el nodo (4, 2) requeriría algunas deducciones adicionales, que van más allá de los métodos desarrollados hasta este punto.

El *método del volumen de control* (también conocido como *método del volumen integral*) ofrece un camino alternativo para la aproximación numérica de las EDP, que es útil en casos como el de la figura 29.13. En la figura 29.12*b*, el método se parece a la aproximación por puntos, donde los puntos se determinan a través del dominio. No obstante, en lugar de aproximar la EDP en un punto, la aproximación se aplica al volumen que rodea el punto. En una malla ortogonal, el volumen está formado por las rectas perpendiculares que pasan por el punto medio de cada línea que une nodos adyacentes. Un balance de calor se obtiene después para cada volumen de manera similar a la ecuación (29.1).

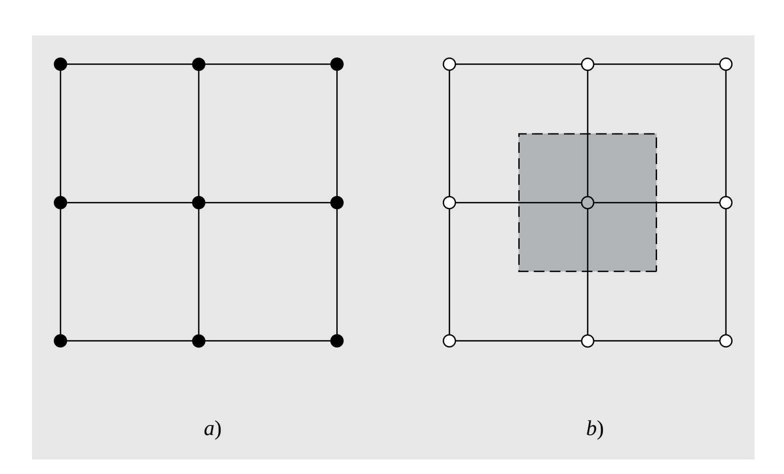


Figura 29.12
Dos perspectivas diferentes para obtener soluciones aproximadas de las EDP: *a*) Método por diferencias finitas punto por punto y *b*) Método del volumen de control.

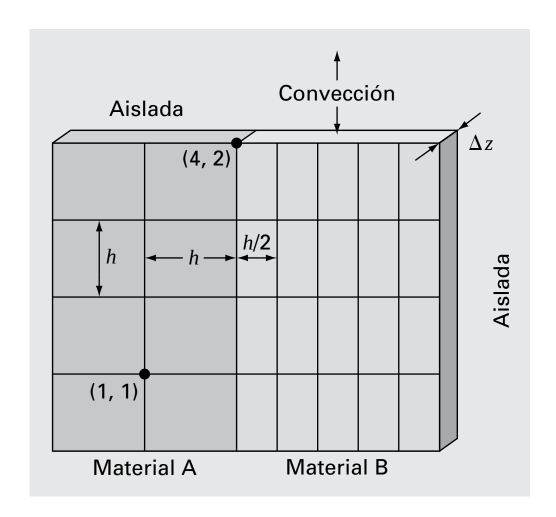


Figura 29.13
Placa caliente con una malla de espaciamientos diferentes, dos materiales y diversas condiciones de frontera.

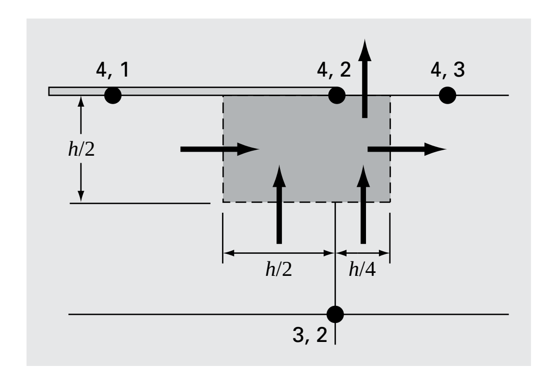


Figura 29.14
Volumen de control para el nodo (4, 2); las flechas indican transferencia de calor a través de las fronteras.

Como ejemplo, aplicaremos el método del volumen de control al nodo (4, 2). Primero, se define el volumen bisecando las rectas que unen los nodos. Como en la figura 29.14, el volumen tiene transferencia de calor por conducción a través de sus fronteras izquierda, derecha e inferior, y la transferencia de calor por convección a través de la mitad de su frontera superior. Observe que la transferencia por la frontera inferior comprende a ambos materiales.

Un balance de calor en estado estacionario para el volumen puede escribirse en términos cualitativos como

$$0 = \begin{pmatrix}\text{conducción}\\ \text{lado izquierdo}\end{pmatrix} - \begin{pmatrix}\text{conducción}\\ \text{lado derecho}\end{pmatrix} + \begin{pmatrix}\text{conducción}\\ \text{inferior material "a"}\end{pmatrix} + \begin{pmatrix}\text{conducción}\\ \text{inferior material "b"}\end{pmatrix} - \begin{pmatrix}\text{convección}\\ \text{superior}\end{pmatrix} \tag{29.27}$$

Ahora el flujo por conducción se representa por la versión en diferencias finitas de la ley de Fourier. Por ejemplo, para el incremento de conducción en el lado izquierdo, sería

$$q = -k_a' \frac{T_{42} - T_{41}}{h}$$

donde las unidades de q son cal/cm²/s. Este flujo se debe multiplicar después por el área transversal a través de la cual entra ($\Delta z \times h/2$), para dar el flujo de calor que entra al volumen por unidad de tiempo,

$$Q = -k_a' \frac{T_{42} - T_{41}}{h} \frac{h}{2} \Delta z$$

donde las unidades de Q son cal/s.

El flujo de calor debido a la convección se formula como sigue

$$q = h_c(T_a - T_{42})$$

donde h_c = un coeficiente por calor de convección [cal/(s · cm² · °C)] y T_a = temperatura del aire (°C). De nuevo, multiplicando por el área adecuada obtenemos la razón del flujo de calor por tiempo,

$$Q = h_c(T_a - T_{42}) \frac{h}{4} \Delta z$$

Las otras transferencias se obtienen de manera similar y se sustituyen en la ecuación (29.27) para dar

$$0 = -k_a' \frac{T_{42} - T_{41}}{h} \frac{h}{2} \Delta z + k_b' \frac{T_{43} - T_{42}}{h/2} \frac{h}{2} \Delta z$$

(conducción lado izquierdo) (conducción lado derecho)

$$-k_a' \frac{T_{42} - T_{32}}{h} \frac{h}{2} \Delta z - k_b' \frac{T_{42} - T_{32}}{h} \frac{h}{4} \Delta z + h_c(T_a - T_{42}) \frac{h}{4} \Delta z$$

$$\begin{pmatrix}\text{Conducción}\\ \text{inferior material "a"}\end{pmatrix} \begin{pmatrix}\text{Conducción}\\ \text{inferior material "b"}\end{pmatrix} (\text{Convección superior})$$

Al sustituir los valores de los parámetros, obtenemos la ecuación final del balance de calor. Por ejemplo, si Δz = 0.5 cm, h = 10 cm, k_a' = 0.3 cal/(s · cm · °C), k_b' = 0.5 cal/(s · cm · °C) y h_c = 0.1 cal/(s · cm² · °C), la ecuación se convierte en

$$0.5875T_{42} - 0.075T_{41} - 0.25T_{43} - 0.1375T_{32} = 2.5$$

Para hacer la ecuación comparable con el laplaciano estándar, ésta se multiplica por 4/0.5875, de modo que el coeficiente del nodo base tenga un coeficiente igual a 4,

$$4T_{42} - 0.510638T_{41} - 1.702128T_{43} - 0.93617T_{32} = 17.02128$$

En los casos estándar vistos hasta ahora, los métodos del volumen de control y de diferencias finitas punto por punto llegan a resultados idénticos. Por ejemplo, en el nodo (1, 1) de la figura 29.13, el balance sería

$$0 = -k_a' \frac{T_{11} - T_{01}}{h} h\, \Delta z + k_a' \frac{T_{21} - T_{11}}{h} h\, \Delta z - k_a' \frac{T_{11} - T_{10}}{h} h\, \Delta z + k_a' \frac{T_{12} - T_{11}}{h} h\, \Delta z$$

que se simplifica al laplaciano estándar,

$$0 = 4T_{11} - T_{01} - T_{21} - T_{12} - T_{10}$$

Veremos otros casos estándar (por ejemplo, la derivada en la condición de frontera) y exploraremos en detalle el método del volumen de control en los problemas del final de este capítulo.

29.5 SOFTWARE PARA RESOLVER ECUACIONES ELÍPTICAS

Modificar un programa computacional para incluir las derivadas en las condiciones de frontera para sistemas rectangulares es una tarea relativamente sencilla. Únicamente consiste en asegurar que se generan ecuaciones adicionales para caracterizar a los nodos frontera donde se especifican las derivadas. Además, hay que modificar el código de tal forma que estas ecuaciones incorporen la derivada como se indica en la ecuación (29.20).

Desarrollar un software general que caracterice los sistemas con fronteras irregulares es mucho más difícil. Por ejemplo, se necesita un algoritmo bastante complicado para modelar la simple junta ilustrada en la figura 29.15. Esto significará dos grandes modificaciones. Primero, se tendrá que desarrollar un esquema para ingresar adecuadamente la configuración de los nodos e identificar los que estén en la frontera. Segundo, se necesitará un algoritmo para generar las ecuaciones simultáneas adecuadas, basándose en la información de entrada. El resultado final es que el software general para resolver las EDP elípticas (y, en general, todas) es relativamente complicado.

Un método utilizado para simplificar este trabajo es proponer una malla muy fina. En tales casos, es frecuente suponer que los nodos cercanos sirven como puntos frontera. De esta manera, el análisis no tiene que considerar los parámetros ponderados de la sección 29.3.2. Aunque esto introduce cierto error, el uso de una malla suficientemente fina puede hacer despreciable la discrepancia resultante. Sin embargo, esto ocasiona una desventaja debido a la carga computacional introducida al aumentar el número de ecuaciones simultáneas.

Como consecuencia de estas consideraciones, el análisis numérico ha desarrollado métodos alternativos que difieren radicalmente de los métodos por diferencias finitas. Aunque estos métodos son conceptualmente más difíciles, pueden implementarse con mayor facilidad para las fronteras irregulares. En el capítulo 31 volveremos a estos métodos. Antes de hacerlo, sin embargo, describiremos los métodos por diferencias finitas en otra categoría de EDP: las ecuaciones parabólicas.

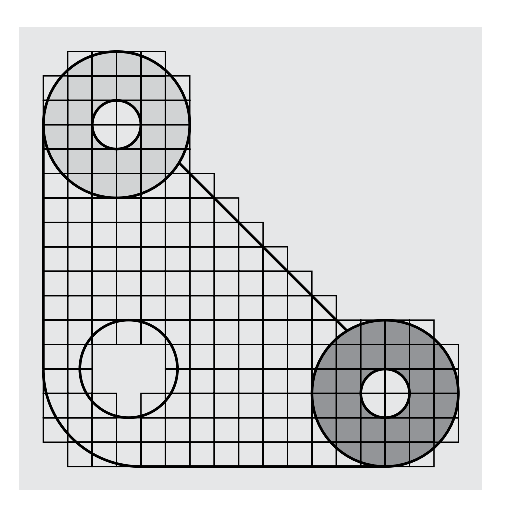


Figura 29.15
Una malla de diferencias finitas sobrepuesta a una junta de forma irregular.

PROBLEMAS

29.1 Use el método de Liebmann para despejar la temperatura de la placa caliente cuadrada de la figura 29.4, pero con la condición de frontera superior aumentada a 150°C, y la frontera izquierda aislada. Use un factor de relajamiento de 1.2 e itere hasta $\varepsilon_s = 1\%$.

29.2 Use el método de Liebmann para resolver cuál sería la temperatura de la placa cuadrada caliente que se ilustra en la figura 29.4, pero con la condición de frontera superior incrementada a 120° y la frontera izquierda disminuida a 60°C. Utilice un factor de relajamiento de 1.2 para iterar a $\varepsilon_s = 1\%$.

29.3 Calcule los flujos para el problema 29.2 con el uso de los parámetros del ejemplo 29.3.

29.4 Repita el ejemplo 29.1, pero emplee 49 nodos interiores (es decir, $\Delta x = \Delta y = 5$ cm).

29.5 Vuelva a hacer el problema 29.4, pero para el caso en que el extremo inferior está aislado.

29.6 Repita los ejemplos 29.1 y 29.3, pero para el caso en que el flujo en el extremo inferior se dirige hacia abajo con un valor de 2 cal/cm^2 · s.

29.7 Repita el ejemplo 29.4 para el caso en que tanto las esquinas inferior izquierda y superior derecha están redondeadas en la misma forma que la esquina inferior izquierda de la figura 29.9. Observe que todas las temperaturas de la frontera en los lados superior y derecho están fijas a 100°C, y todas las de los lados inferior e izquierdo lo están a 50°C.

29.8 Con excepción de las condiciones de frontera, la placa de la figura P29.8 tiene las mismas características que la que se usó en los ejemplos 29.1 a 29.3. Para dicha placa, simule tanto las temperaturas como los flujos.

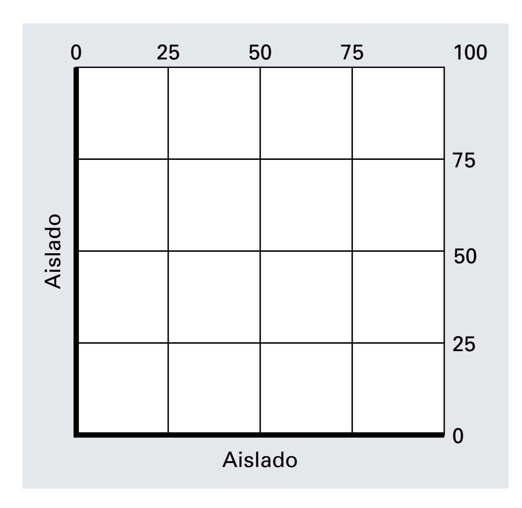


Figura P29.8

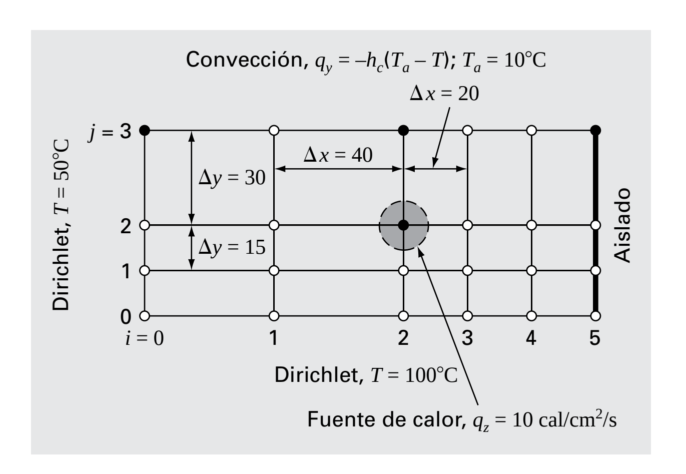


Figura P29.9

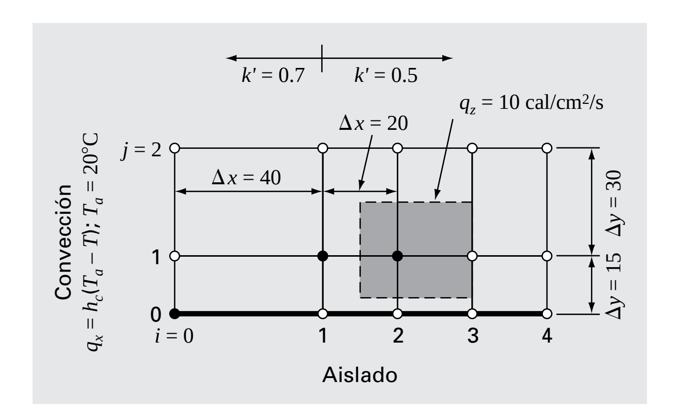


Figura P29.10

29.9 Escriba ecuaciones para los nodos resaltados en la malla que se ilustra en la figura P29.9. Observe que todas las unidades son del sistema cgs. El coeficiente de conductividad térmica para la placa es de 0.75 cal/(s · cm · °C), el coeficiente de convección es $h_c = 0.015$ cal/(cm^2 · C · s), y el espesor de la placa es de 0.5 cm.

29.10 Escriba ecuaciones para los nodos resaltados en la malla que aparece en la figura P29.10. Observe que todas las unidades son del sistema cgs. El coeficiente de convección es $h_c = 0.01$ cal/(cm^2 · C · s), y el espesor de la placa es de 2 cm.

29.11 Aplique el enfoque del volumen de control para desarrollar la ecuación para el nodo (0, *j*) de la figura 29.7.

29.12 Deduzca una ecuación como la ecuación (29.26) en el caso donde θ es mayor a 45° para la figura 29.11.

29.13 Desarrolle un programa de computadora amigable para el usuario para implantar el método de Liebmann para una placa rectangular con condiciones de frontera de Dirichlet. Diseñe el programa de modo que calcule tanto la temperatura como el flujo. Pruebe el programa con la duplicación de los resultados de los ejemplos 29.1 y 29.2.

29.14 Emplee el programa del problema 29.13 para resolver los problemas 29.2 y 29.3.

29.15 Utilice el programa del problema 29.13 para resolver el problema 29.4.

29.16 Emplee el enfoque del volumen de control y obtenga la ecuación de nodo para el nodo (2, 2) de la figura 29.13, e incluya una fuente de calor en este punto. Utilice los valores siguientes para las constantes: $\Delta z = 0.25$ cm, $h = 10$ cm, $k_A = 0.25$ W/cm · C y $k_B = 0.45$ W/cm · C. La fuente calorífica sólo proviene del material *A* a una tasa de 6 W/cm^3.

29.17 Calcule el flujo de calor (W/cm²) en el nodo (2, 2) de la figura 29.13, con aproximaciones por diferencias finitas para los gradientes de temperatura en dicho nodo. Calcule el flujo en dirección horizontal en los materiales *A* y *B* y determine si los dos flujos deben ser iguales. Asimismo, calcule el flujo vertical en los materiales *A* y *B*. ¿Deben ser iguales estos dos flujos? Utilice los valores siguientes para las constantes: $\Delta z = 0.5$ cm, $h = 10$ cm, $k_A = 0.25$ W/cm · C, $k_B = 0.45$ W/cm · C, y temperaturas en los nodos: $T_{22} = 51.6$°C, $T_{21} = 74.2$°C, $T_{23} = 45.3$°C, $T_{32} = 38.6$°C y $T_{12} = 87.4$°C.

29.18 Calcule la distribución de temperatura para la placa en forma de L de la figura P29.18.

29.19 La ecuación de Poisson se puede escribir en tres dimensiones como

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} = f(x, y, z)$$

Despeje la distribución de temperatura dentro de un cubo unitario (1 × 1) con condiciones de frontera cero y $f = -10$. Utilice $\Delta x = \Delta y = \Delta z = 1/6$.

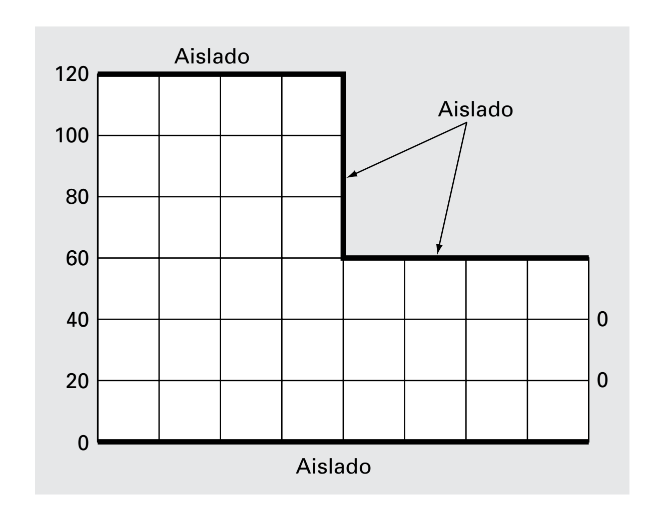


Figura P29.18

CAPÍTULO 30

Diferencias finitas: ecuaciones parabólicas

En el capítulo anterior tratamos las EDP en estado estacionario. Ahora veremos las ecuaciones parabólicas que se emplean para caracterizar problemas que varían con el tiempo. En la última parte de este capítulo, ilustraremos cómo se desarrollan estos problemas en dos dimensiones espaciales para la placa calentada. Antes, mostraremos cómo se aborda el caso unidimensional más sencillo.

30.1 LA ECUACIÓN DE CONDUCCIÓN DE CALOR

De manera similar a la deducción de la ecuación de Laplace [ecuación (29.6)], se puede utilizar la conservación del calor para desarrollar un balance de calor del elemento diferencial, en la barra larga, delgada y aislada que se muestra en la figura 30.1. Sin embargo, en lugar de examinar el caso en estado estacionario, este balance también considera la cantidad de calor que se almacena en el elemento en un periodo Δt. El balance tiene la forma, *entradas – salidas = acumulación*, o

$$q(x)\,\Delta y\,\Delta z\,\Delta t - q(x+\Delta x)\,\Delta y\,\Delta z\,\Delta t = \Delta x\,\Delta y\,\Delta z\rho C\,\Delta T$$

Dividiendo entre el volumen del elemento ($= \Delta x\,\Delta y\,\Delta z$) y entre Δt se obtiene

$$\frac{q(x)-q(x+\Delta x)}{\Delta x} = \rho C \frac{\Delta T}{\Delta t}$$

Tomando el límite se llega a

$$-\frac{\partial q}{\partial x} = \rho C \frac{\partial T}{\partial t}$$

Sustituyendo la ley de Fourier para la conducción del calor [ecuación (29.4)] se obtiene

$$k\frac{\partial^2 T}{\partial x^2} = \frac{\partial T}{\partial t} \tag{30.1}$$

que es la *ecuación de conducción del calor.*

De la misma manera que con las EDP elípticas, las ecuaciones parabólicas se resuelven sustituyendo las derivadas parciales por diferencias divididas finitas. Sin embargo, a diferencia de las EDP elípticas, debemos considerar cambios tanto en el tiempo como en el espacio. Mientras que las ecuaciones elípticas están acotadas en todas las dimensiones, las EDP parabólicas están temporalmente abiertas en los extremos (figura 30.2). Debido a su naturaleza variable en el tiempo, las soluciones de estas ecuaciones presentan problemas nuevos, notablemente relacionados con

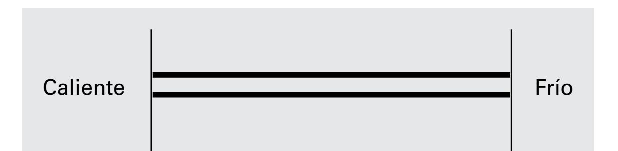


Figura 30.1
Una barra delgada y aislada en todos los puntos excepto en sus extremos.

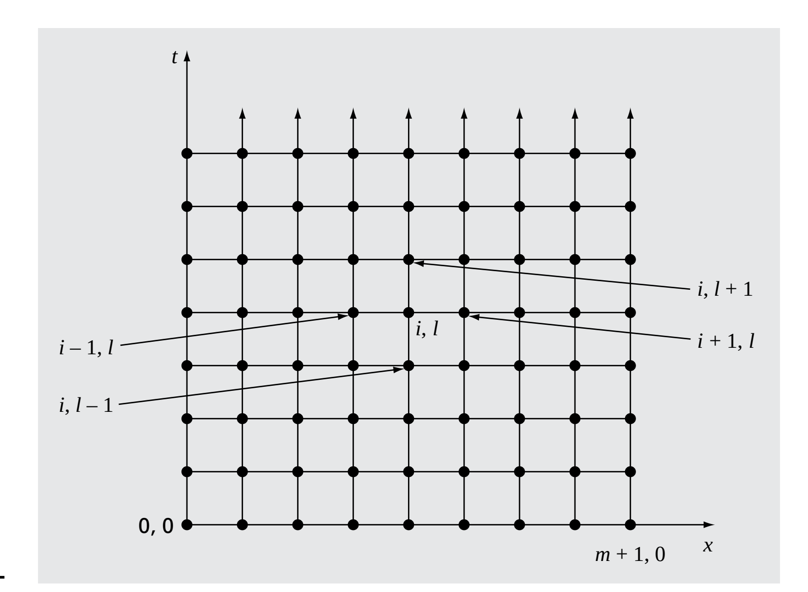


Figura 30.2
Una malla utilizada para la solución por diferencias finitas de las EDP parabólicas con dos variables independientes, por ejemplo la ecuación de conducción del calor. Observe cómo, a diferencia de la figura 29.3, la malla está abierta en los extremos de la dimensión temporal.

la estabilidad. Éste y otros aspectos de las EDP parabólicas se examinarán en las secciones siguientes, donde presentamos fundamentalmente dos métodos de solución: los esquemas explícitos y los implícitos.

30.2 MÉTODOS EXPLÍCITOS

La ecuación de conducción del calor requiere aproximaciones de la segunda derivada en el espacio, y de la primera derivada en el tiempo. La segunda derivada se representa, de la misma manera que la ecuación de Laplace, mediante una diferencia dividida finita centrada:

$$\frac{\partial^2 T}{\partial x^2} = \frac{T_{i+1}^l - 2T_i^l + T_{i-1}^l}{\Delta x^2} \tag{30.2}$$

que tiene un error (recuerde la figura 23.3) de $O[(\Delta x)^2]$. Observe que el ligero cambio en la notación de los superíndices se utiliza para denotar tiempo. Esto se hace para que un segundo subíndice pueda usarse para designar una segunda dimensión espacial cuando el método se extiende a dos dimensiones espaciales.

Una diferencia dividida finita hacia adelante sirve para aproximar a la derivada con respecto al tiempo

$$\frac{\partial T}{\partial t} = \frac{T_i^{l+1} - T_i^l}{\Delta t} \tag{30.3}$$

la cual tiene un error (recuerde la figura 23.1) de $O(\Delta t)$.

Sustituyendo las ecuaciones (30.2) y (30.3) en la ecuación (30.1), se obtiene

$$k\frac{T_{i+1}^l - 2T_i^l + T_{i-1}^l}{(\Delta x)^2} = \frac{T_i^{l+1} - T_i^l}{\Delta t} \tag{30.4}$$

de donde resulta

$$T_i^{l+1} = T_i^l + \lambda(T_{i+1}^l - 2T_i^l + T_{i-1}^l) \tag{30.5}$$

donde $\lambda = k\,\Delta t/(\Delta x)^2$.

Esta ecuación se puede escribir para todos los nodos interiores de la barra. Dicha ecuación proporciona un medio explícito para calcular los valores en cada nodo para un tiempo posterior,

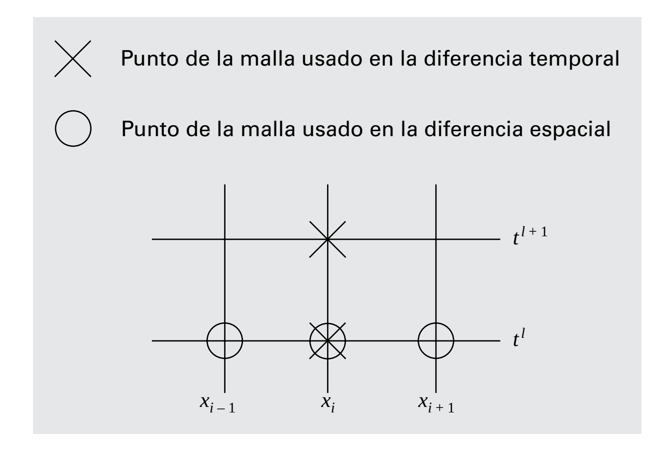


Figura 30.3
Molécula computacional para la forma explícita.

basándose en los valores presentes del nodo y de sus vecinos. Observe que este método es una manifestación del método de Euler para resolver sistemas de EDO. Es decir, si conocemos la distribución de temperatura como una función de la posición en un tiempo inicial, es posible calcular la distribución en un tiempo futuro, basada en la ecuación (30.5).

Una molécula computacional para el método explícito se representa en la figura 30.3; ahí se muestran los nodos que constituyen las aproximaciones espacial y temporal. Esta molécula se compara con otras de este capítulo para ilustrar las diferencias entre los métodos.

EJEMPLO 30.1 Solución explícita para la ecuación de conducción de calor unidimensional

Planteamiento del problema Con el método explícito calcule la distribución de temperatura en una barra larga y delgada que tiene una longitud de 10 cm y los siguientes valores: $k' = 0.49$ cal/(s · cm · °C), $\Delta x = 2$ cm y $\Delta t = 0.1$ s. En $t = 0$, la temperatura de la barra es cero, y las condiciones de frontera se fijan para todos los tiempos en $T(0) = 100$°C y $T(10) = 50$°C. Considere que la barra es de aluminio con $C = 0.2174$ cal/(g · °C) y $\rho = 2.7$ g/cm^3. Por lo tanto, $k = 0.49/(2.7 \cdot 0.2174) = 0.835$ cm^2/s y $\lambda = 0.835(0.1)/(2)^2 = 0.020875$.

Solución Aplicando la ecuación (30.5) se obtiene el siguiente valor en $t = 0.1$ s para el nodo en $x = 2$ cm:

$$T_1^1 = 0 + 0.020875[0 - 2(0) + 100] = 2.0875$$

En los otros puntos interiores, $x = 4$, 6 y 8 cm, los resultados son

$$T_2^1 = 0 + 0.020875[0 - 2(0) + 0] = 0$$
$$T_3^1 = 0 + 0.020875[0 - 2(0) + 0] = 0$$
$$T_4^1 = 0 + 0.020875[50 - 2(0) + 0] = 1.0438$$

En $t = 0.2$ s, los valores obtenidos para los cuatro nodos interiores son

$$T_1^2 = 2.0875 + 0.020875[0 - 2(2.0875) + 100] = 4.0878$$
$$T_2^2 = 0 + 0.020875[0 - 2(0) + 2.0875] = 0.043577$$
$$T_3^2 = 0 + 0.020875[1.0438 - 2(0) + 0] = 0.021788$$
$$T_4^2 = 1.0438 + 0.020875[50 - 2(1.0438) + 0] = 2.0439$$

El cálculo continúa y los resultados en intervalos de 3 segundos se ilustran en la figura 30.4. El aumento general de la temperatura con el tiempo indica que el cálculo capta la difusión del calor desde las fronteras del interior de la barra.

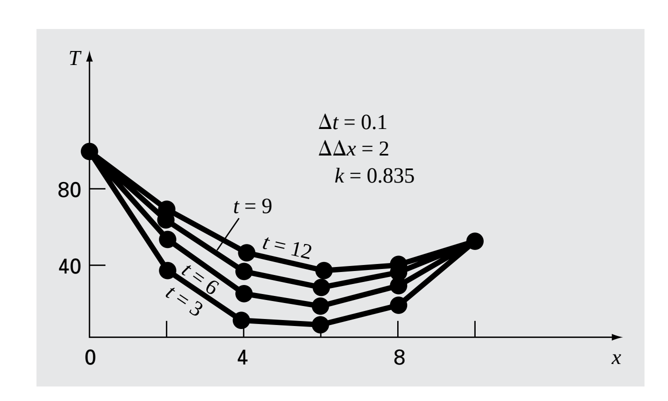


Figura 30.4
Distribución de temperatura en una barra larga y delgada, calculada con el método explícito que se describe en la sección 30.2.

30.2.1 Convergencia y estabilidad

Convergencia significa que conforme Δx y Δt tiendan a cero, los resultados de la técnica por diferencias finitas se aproximarán a la solución verdadera. *Estabilidad* significa que los errores en cualquier etapa del cálculo no se amplifican, sino que se atenúan conforme avanza el cálculo. Se puede demostrar (véase Carnahan y cols., 1969) que el método explícito es convergente y estable si $\lambda \leq 1/2$, o

$$\Delta t \leq \frac{1}{2}\frac{\Delta x^2}{k} \tag{30.6}$$

Además, se debe observar que cuando $\lambda \leq 1/2$ se puede tener como resultado una solución donde los errores no crecen, sino que oscilan. Haciendo $\lambda \leq 1/4$ asegura que la solución no oscilará. También se sabe que con $\lambda = 1/6$ se tiende a minimizar los errores por truncamiento (véase Carnahan y cols., 1969).

La figura 30.5 es un ejemplo de inestabilidad causada al violar la ecuación (30.6). Esta gráfica es para el mismo caso del ejemplo 30.1, pero con $\lambda = 0.735$, que es considerablemente mayor que 0.5. Como se advierte en la figura 30.5, la solución experimenta en forma progresiva mayores oscilaciones. Esta situación continuará conforme el cálculo continúa.

Aunque al satisfacer la ecuación (30.6) se disminuirían las inestabilidades del tipo mostrado en la figura 30.5, también impone fuertes limitaciones al método explícito. Por ejemplo, suponga que Δx se reduce a la mitad para mejorar la aproximación de la segunda derivada espacial. De acuerdo con la ecuación (30.6) el tamaño de paso para el tiempo debe reducirse a un cuarto para mantener la convergencia y la estabilidad. Así, para realizar cálculos comparables, el número de pasos debe aumentar por un factor de 4. Es más, el cálculo para cada uno de estos tamaños de paso del tiempo tomará el doble de tiempo, ya que al dividir Δx a la mitad se duplica el número total de nodos para los cuales hay que aplicar las ecuaciones. En consecuencia, en el caso unidimensional, reducir Δx a la mitad da como resultado un aumento de ocho veces en el número de cálculos. Así, la carga computacional puede resultar tan grande que impida alcanzar una exactitud aceptable. Como describiremos en breve, hay otras técnicas que no adolecen de limitantes tan severas.

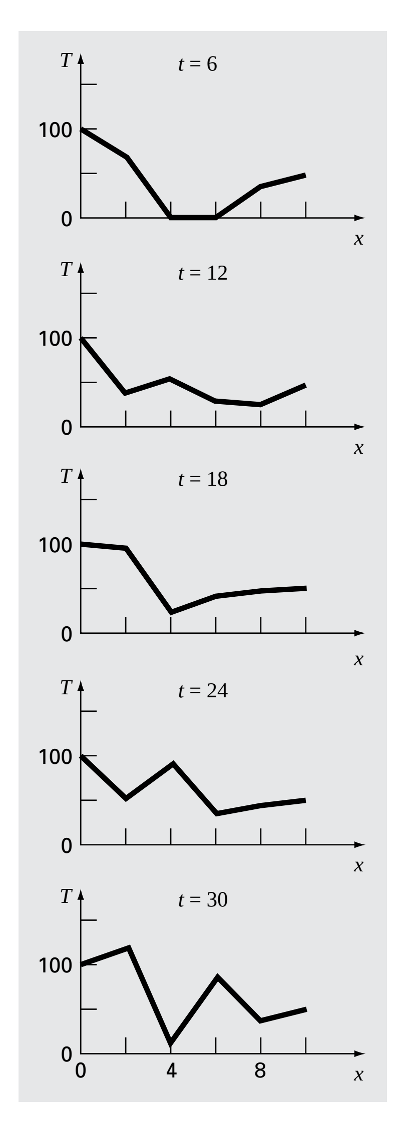


Figura 30.5
Ilustración de la inestabilidad. Solución del ejemplo 30.1, pero con $\lambda = 0.735$.

30.2.2 La derivada en las condiciones de frontera

Como en el caso de las EDP elípticas (recuerde la sección 29.3.1), la derivada en las condiciones de frontera se puede incorporar fácilmente a las ecuaciones parabólicas. Para una barra unidimensional, se necesita agregar dos ecuaciones para caracterizar el balance de calor en los nodos extremos. Por ejemplo, el nodo del extremo izquierdo ($i = 0$) se representará por

$$T_0^{l+1} = T_0^l + \lambda(T_1^l - 2T_0^l + T_{-1}^l)$$

Así, se introdujo un punto imaginario en $i = -1$ (recuerde la figura 29.7). Sin embargo, como en el caso elíptico, este punto ofrece un medio para incorporar en el análisis la derivada en las condiciones de frontera. El problema 30.2, que está al final de este capítulo, se ocupa de este ejercicio.

30.2.3 Aproximaciones temporales de orden superior

La idea general de volver a expresar la EDP como un sistema de EDO se denomina *método de líneas*. En efecto, una manera de mejorar el método de Euler usado antes, sería emplear un esquema de integración más exacto para resolver las EDO. Por ejemplo, el método de Heun puede utilizarse para obtener una exactitud temporal de segundo orden. Esta aproximación se conoce como *método de MacCormack*. Éste y otros métodos explícitos mejorados se analizan en obras como, por ejemplo, la de Hoffmann (1992).

30.3 UN MÉTODO IMPLÍCITO SIMPLE

Como ya se indicó, las formulaciones explícitas por diferencias finitas tienen problemas relacionados con la estabilidad. Además, como se ilustra en la figura 30.6, excluyen información de importancia para la solución. Los métodos implícitos superan ambas dificultades a expensas de utilizar algoritmos un poco más complicados.

La diferencia fundamental entre los métodos explícitos y los implícitos se ilustra en la figura 30.7. En la forma explícita, aproximamos la derivada espacial para un nivel de tiempo l (figura 30.7*a*). Recuerde que cuando sustituimos esta aproximación en la ecuación diferencial parcial, obtuvimos una ecuación en diferencias (30.4) con una sola incógnita T_i^{l+1}. Así, podemos despejar "explícitamente" esta incógnita como en la ecuación (30.5).

En los métodos implícitos, la derivada espacial se aproxima en un nivel de tiempo posterior $l + 1$. Por ejemplo, la segunda derivada se aproximará mediante (figura 30.7*b*)

$$\frac{\partial^2 T}{\partial x^2} \cong \frac{T_{i+1}^{l+1} - 2T_i^{l+1} + T_{i-1}^{l+1}}{(\Delta x)^2} \tag{30.7}$$

que tiene una exactitud de segundo orden. Cuando esta relación se sustituye en la EDP original, la ecuación en diferencias resultante contiene varias incógnitas. Así, no puede resolverse explícitamente mediante simples manipulaciones algebraicas, como se hizo al pasar de la ecuación (30.4) a la (30.5). En lugar de esto, el sistema completo de ecuaciones debe resolverse simultáneamente. Esto es posible debido a que, junto con las condiciones de frontera, las formulaciones implícitas dan como resultado un conjunto de ecuaciones lineales algebraicas con el mismo número de incógnitas. Por lo tanto, el método se reduce a la solución de un sistema de ecuaciones simultáneas en cada punto en el tiempo.

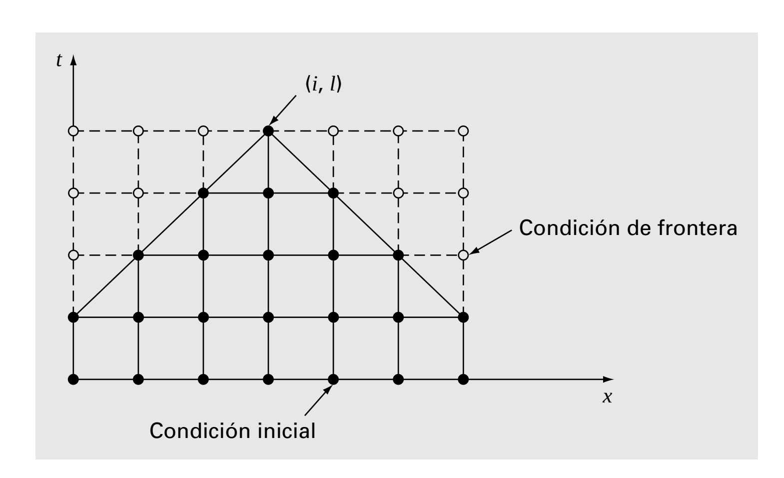


Figura 30.6
Representación del efecto de los otros nodos sobre la aproximación por diferencias finitas en el nodo (*i*, *l*) usando un esquema explícito por diferencias finitas. Los nodos marcados tienen una influencia sobre (*i*, *l*); en tanto que los nodos no marcados, que en realidad afectan a (*i*, *l*), se han excluido.

Para ilustrar cómo hacer lo anterior, sustituimos las ecuaciones (30.3) y (30.7) en la ecuación (30.1), para obtener

$$k\frac{T_{i+1}^{l+1} - 2T_i^{l+1} + T_{i-1}^{l+1}}{(\Delta x)^2} = \frac{T_i^{l+1} - T_i^l}{\Delta t}$$

que se expresa como

$$-\lambda T_{i-1}^{l+1} + (1 + 2\lambda)T_i^{l+1} - \lambda T_{i+1}^{l+1} = T_i^l \tag{30.8}$$

donde $\lambda = k\,\Delta t/(\Delta x)^2$. Esta ecuación se aplica a todos los nodos, excepto al primero y al último de los nodos interiores, los cuales deben modificarse para considerar las condiciones de frontera. En el caso donde están dados los niveles de temperatura en los extremos de la barra, la condición de frontera en el extremo izquierdo de la barra ($i = 0$) se expresa como

$$T_0^{l+1} = f_0(t^{l+1}) \tag{30.9}$$

donde $f_0(t^{l+1})$ = una función que describe cómo cambia con el tiempo la temperatura de la frontera. Sustituyen-

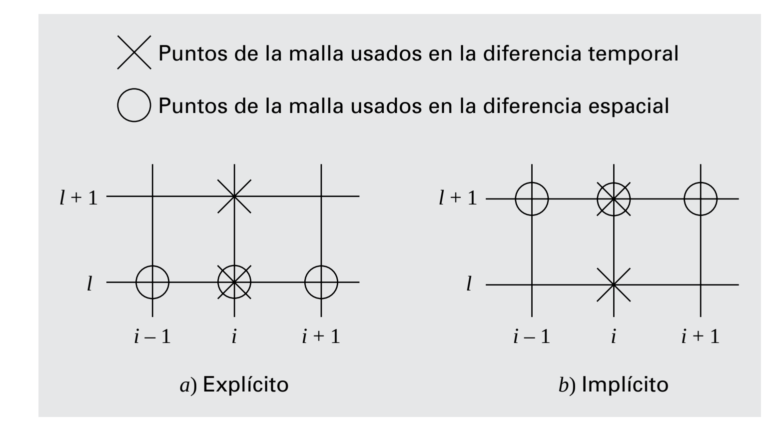


Figura 30.7
Moléculas computacionales que demuestran las diferencias fundamentales entre los métodos *a*) explícito y *b*) implícito.

do la ecuación (30.9) en la ecuación (30.8), se obtiene la ecuación en diferencias para el primer nodo interior ($i = 1$):

$$(1 + 2\lambda)T_1^{l+1} - \lambda T_2^{l+1} = T_1^l + \lambda f_0(t^{l+1}) \tag{30.10}$$

De manera similar, para el último nodo interior ($i = m$),

$$-\lambda T_{m-1}^{l+1} + (1 + 2\lambda)T_m^{l+1} = T_m^l + \lambda f_{m+1}(t^{l+1}) \tag{30.11}$$

donde $f_{m+1}(t^{l+1})$ describe los cambios específicos de temperatura en el extremo derecho de la barra ($i = m + 1$).

Cuando se escriben las ecuaciones (30.8), (30.10) y (30.11) para todos los nodos interiores, el conjunto resultante de m ecuaciones algebraicas lineales tiene m incógnitas. Además, el método tiene la ventaja de que el sistema es tridiagonal. Así, es posible utilizar los algoritmos de solución extremadamente eficientes (recuerde la sección 11.1.1) disponibles para sistemas tridiagonales.

EJEMPLO 30.2 Solución implícita simple de la ecuación de conducción del calor

Planteamiento del problema Con la aproximación por diferencias finitas implícita simple resuelva el problema del ejemplo 30.1.

Solución Para la barra del ejemplo 30.1, $\lambda = 0.020875$. Por lo tanto, en $t = 0$, la ecuación (30.10) para el primer nodo interior se escribe como

$$1.04175T_1^1 - 0.020875T_2^1 = 0 + 0.020875(100) \quad \text{o} \quad 1.04175T_1^1 - 0.020875T_2^1 = 2.0875$$

De manera similar, las ecuaciones (30.8) y (30.11) pueden aplicarse a los otros nodos interiores. Esto nos conduce al siguiente sistema de ecuaciones simultáneas:

$$\begin{bmatrix} 1.04175 & -0.020875 & & \\ -0.020875 & 1.04175 & -0.020875 & \\ & -0.020875 & 1.04175 & -0.020875 \\ & & -0.020875 & 1.04175 \end{bmatrix} \begin{Bmatrix} T_1^1 \\ T_2^1 \\ T_3^1 \\ T_4^1 \end{Bmatrix} = \begin{Bmatrix} 2.0875 \\ 0 \\ 0 \\ 1.04375 \end{Bmatrix}$$

que se resuelve para la temperatura en $t = 0.1$ s:

$$T_1^1 = 2.0047$$
$$T_2^1 = 0.0406$$
$$T_3^1 = 0.0209$$
$$T_4^1 = 1.0023$$

Observe cómo, a diferencia del ejemplo 30.1, todos los puntos se han modificado de la condición inicial durante el primer paso de tiempo.

Al resolverlas para temperaturas en $t = 0.2$, el vector del lado derecho debe modificarse considerando los resultados del primer paso, así

$$\begin{Bmatrix} 4.09215 \\ 0.04059 \\ 0.02090 \\ 2.04069 \end{Bmatrix}$$

Entonces, de las ecuaciones simultáneas se obtienen las temperaturas en $t = 0.2$ s:

$$T_1^2 = 3.9305$$
$$T_2^2 = 0.1190$$
$$T_3^2 = 0.0618$$
$$T_4^2 = 1.9653$$

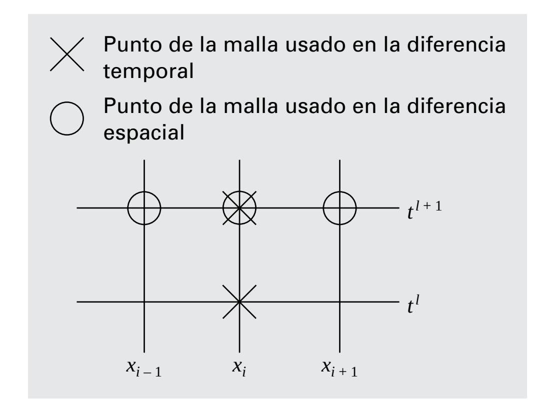


Figura 30.8
Una molécula computacional para el método implícito simple.

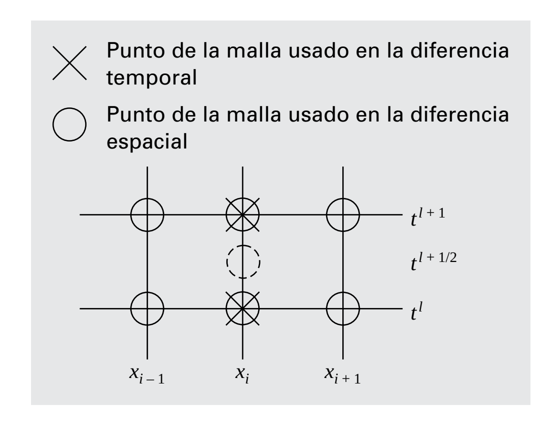


Figura 30.9
Molécula computacional para el método de Crank-Nicolson.

Mientras que el método implícito descrito es estable y convergente, es deficiente en el sentido de que la aproximación en diferencias temporal tiene una exactitud de primer orden; en tanto que la aproximación en diferencias espacial tiene exactitud de segundo orden (figura 30.8). En la siguiente sección presentaremos un método implícito alternativo que resuelve esta situación.

Antes de continuar, hay que mencionar que, aunque el método implícito simple es incondicionalmente estable, hay un límite de exactitud para el uso de pasos de tiempo grandes. En consecuencia, no es mucho más eficiente que los métodos explícitos para la mayoría de los problemas variables en el tiempo.

Donde se nota esto es en los problemas en estado estacionario. Del capítulo 29 recuerde que una forma de Gauss-Seidel (método de Liebmann) se utiliza para obtener soluciones para estado estacionario de las ecuaciones elípticas. Un método alternativo será correr una solución variable en el tiempo hasta que alcance un estado estacionario. En estos casos, debido a que los resultados intermedios inexactos no son un problema, los métodos implícitos permiten emplear grandes pasos de tiempo, y así generar resultados a estado estacionario de manera eficiente.

30.4 EL MÉTODO DE CRANK-NICOLSON

El *método de Crank-Nicolson* ofrece un esquema implícito alternativo que tiene una exactitud de segundo orden, tanto para el espacio como para el tiempo. Para alcanzar tal exactitud, se desarrollan aproximaciones por diferencias en el punto medio del incremento del tiempo (figura 30.9). Entonces, la primera derivada temporal se aproxima en $t^{l+1/2}$ por

$$\frac{\partial T}{\partial t} \cong \frac{T_i^{l+1} - T_i^l}{\Delta t} \tag{30.12}$$

La segunda derivada en el espacio puede determinarse en el punto medio promediando las aproximaciones por diferencias al principio (t^l) y al final (t^{l+1}) del incremento del tiempo

$$\frac{\partial^2 T}{\partial x^2} \cong \frac{1}{2}\left[\frac{T_{i+1}^l - 2T_i^l + T_{i-1}^l}{(\Delta x)^2} + \frac{T_{i+1}^{l+1} - 2T_i^{l+1} + T_{i-1}^{l+1}}{(\Delta x)^2}\right] \tag{30.13}$$

Sustituyendo las ecuaciones (30.12) y (30.13) en la ecuación (30.1) y reagrupando términos, se obtiene

$$-\lambda T_{i-1}^{l+1} + 2(1+\lambda)T_i^{l+1} - \lambda T_{i-1}^{l+1} = \lambda T_{i-1}^l + 2(1-\lambda)T_i^l + \lambda T_{i+1}^l \tag{30.14}$$

donde $\lambda = k\,\Delta t/(\Delta x)^2$. Como en el caso del método implícito simple, se determinan las condiciones de frontera $T_0^{l+1} = f_0(t^{l+1})$ y $T_{m+1}^{l+1} = f_{m+1}(t^{l+1})$ para obtener versiones de la ecuación (30.14) para los nodos interiores primero y último. Para el primer nodo interior

$$2(1+\lambda)T_1^{l+1} - \lambda T_2^{l+1} = \lambda f_0(t^l) + 2(1-\lambda)T_1^l + \lambda T_2^l + \lambda f_0(t^{l+1}) \tag{30.15}$$

y, para el último nodo interior,

$$-\lambda T_{m+1}^{l+1} + 2(1+\lambda)T_m^{l+1} = \lambda f_{m+1}(t^l) + 2(1-\lambda)T_m^l + \lambda T_{m-1}^l + \lambda f_{m+1}(t^{l+1}) \tag{30.16}$$

Aunque las ecuaciones (30.14) a (30.16) son ligeramente más complicadas que las ecuaciones (30.8), (30.10) y (30.11), también son tridiagonales y, por lo tanto, se resuelve de manera eficiente.

EJEMPLO 30.3 Solución de Crank-Nicolson para la ecuación de conducción del calor

Planteamiento del problema Con el método de Crank-Nicolson resuelva el mismo problema que en los ejemplos 30.1 y 30.2.

Solución Las ecuaciones (30.14) a (30.16) se utilizan para generar el siguiente sistema de ecuaciones tridiagonal:

$$\begin{bmatrix} 2.04175 & -0.020875 & & \\ -0.20875 & 2.04175 & -0.020875 & \\ & -0.020875 & 2.04175 & -0.020875 \\ & & -0.020875 & 2.04175 \end{bmatrix} \begin{Bmatrix} T_1^1 \\ T_2^1 \\ T_3^1 \\ T_4^1 \end{Bmatrix} = \begin{Bmatrix} 4.175 \\ 0 \\ 0 \\ 2.0875 \end{Bmatrix}$$

de donde se obtienen las temperaturas en $t = 0.1$ s:

$$T_1^1 = 2.0450$$
$$T_2^1 = 0.0210$$
$$T_3^1 = 0.0107$$
$$T_4^1 = 1.0225$$

Para obtener las temperaturas en $t = 0.2$ s, el vector del lado derecho debe modificarse

$$\begin{Bmatrix} 8.1801 \\ 0.0841 \\ 0.0427 \\ 4.0901 \end{Bmatrix}$$

De las ecuaciones simultáneas se obtiene

$$T_1^2 = 4.0073$$
$$T_2^2 = 0.0826$$
$$T_3^2 = 0.0422$$
$$T_4^2 = 2.0036$$

30.4.1 Comparación de los métodos unidimensionales

La ecuación (30.1) se puede resolver en forma analítica. Por ejemplo, hay una solución para el caso donde la temperatura de la barra es inicialmente cero. En $t = 0$, la condición de frontera en $x = L$ se eleva instantáneamente a un nivel constante de $\overline{T}$, mientras que $T(0)$ se mantiene en cero. En este caso, la temperatura se calcula por

$$T = \overline{T}\left[\frac{x}{L} + \sum_{n=0}^{\infty} \frac{2}{n\pi}(-1)^n \operatorname{sen}\left(\frac{nx}{L}\right)\exp\left(\frac{-n^2\pi^2 kt}{L^2}\right)\right] \tag{30.17}$$

donde L = longitud total de la barra. Esta ecuación es útil para calcular la evolución de la distribución de temperaturas para cada condición de frontera. Entonces, la solución total se determina por superposición.

EJEMPLO 30.4 Comparación de las soluciones numéricas y analíticas

Planteamiento del problema Compare la solución analítica de la ecuación (30.17) con los resultados numéricos obtenidos con las técnicas explícita, implícita simple y de Crank-Nicolson. Realice esta comparación con la barra empleada en los ejemplos 30.1, 30.2 y 30.3.

Solución Recuerde de los ejemplos anteriores que $k = 0.835$ cm^2/s, $L = 10$ cm y $\Delta x = 2$ cm. En este caso, se utiliza la ecuación (30.17) para predecir que la temperatura en $x = 2$ cm y $t = 10$ s será igual a 64.8018. En la tabla 30.1 se presentan predicciones numéricas para $T(2, 10)$. Observe que se ha empleado un tamaño de paso para el tiempo. Estos resultados indican varias propiedades de los métodos numéricos. Primero, se observa que el método explícito es inestable para λ alta. Dicha inestabilidad no se manifiesta en ningún método implícito. Segundo, el método de Crank-Nicolson converge más rápidamente conforme λ decrece, y proporciona resultados de exactitud moderada aun cuando λ sea relativamente alta. Estos resultados eran de esperarse ya que Crank-Nicolson tiene una exactitud de segundo orden con respecto a ambas variables independientes. Por último, observe que conforme λ decrece, los métodos parecen converger a un valor de 64.73, que es diferente del resultado analítico de 64.80. Esto no debe sorprender, ya que se ha usado un valor fijo de $\Delta x = 2$ para caracterizar la dimensión x. Si tanto Δx como Δt disminuyeran conforme λ decrece (es decir, si se usaran más segmentos espaciales), la solución numérica se acercará más al resultado analítico.

Tabla 30.1 Comparación de tres métodos para la solución de una EDP parabólica: la barra calentada. Los resultados mostrados corresponden a la temperatura en $t = 10$ s en $x = 2$ cm para la barra de los ejemplos 30.1 a 30.3. Observe que la solución analítica es $T(2, 10) = 64.8018$.

Δt	λ	Explícito	Implícito	Crank-Nicolson
10	2.0875	208.75	53.01	79.77
5	1.04375	–9.13	58.49	64.79
2	0.4175	67.12	62.22	64.87
1	0.20875	65.91	63.49	64.77
0.5	0.104375	65.33	64.12	64.74
0.2	0.04175	64.97	64.49	64.73

El método de Crank-Nicolson se emplea con frecuencia para resolver EDP parabólicas en una dimensión espacial. Las ventajas del método se aprecian cuando se presentan problemas más complicados, como aquellos en los que se tienen mallas irregularmente espaciadas. Tal espaciado no uniforme a menudo es ventajoso cuando se tiene un conocimiento previo de que la solución varía rápidamente en porciones locales del sistema. Análisis de tales aplicaciones y del método de Crank-Nicolson se encuentran en diferentes fuentes (Ferziger, 1981; Lapidus y Pinder, 1981; Hoffman, 1992).

30.5 ECUACIONES PARABÓLICAS EN DOS DIMENSIONES ESPACIALES

La ecuación de conducción del calor se puede aplicar a más de una dimensión espacial. Para dos dimensiones, su forma es

$$\frac{\partial T}{\partial t} = k\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2}\right) \tag{30.18}$$

Una aplicación de esta ecuación consiste en modelar la distribución de temperatura sobre la superficie de una placa calentada. Sin embargo, más que caracterizar su distribución en estado estacionario, como se hizo en el capítulo 29, la ecuación (30.18) ofrece un medio para calcular la distribución de temperatura de la placa conforme cambia con el tiempo.

30.5.1 Esquemas explícito o implícito estándar

Es posible obtener una solución explícita sustituyendo en la ecuación (30.18) las aproximaciones por diferencias finitas de la forma de las ecuaciones (30.2) y (30.3). Sin embargo, como en el caso unidimensional, este método está limitado por un estricto criterio de estabilidad. En el caso bidimensional, el criterio es

$$\Delta t \leq \frac{1}{8}\frac{(\Delta x)^2 + (\Delta y)^2}{k}$$

Así, para una malla uniforme ($\Delta x = \Delta y$), $\lambda = k\,\Delta t/(\Delta x)^2$ debe ser menor o igual que 1/4. En consecuencia, si se reduce a la mitad, el tamaño del paso, se cuadruplica el número de nodos y aumenta el trabajo computacional en un factor de 16.

Como en el caso de sistemas unidimensionales, las técnicas implícitas ofrecen alternativas que garantizan estabilidad. Sin embargo, la aplicación directa de los métodos implícitos, como la técnica de Crank-Nicolson, nos lleva a la solución de $m \times n$ ecuaciones simultáneas. Además, cuando se aplican para dos o tres dimensiones espaciales, estas ecuaciones pierden la valiosa propiedad de ser tridiagonales. De esta manera, el almacenamiento de la matriz y el tiempo de cálculo llegan a ser extremadamente grandes. El método descrito en la siguiente sección ofrece una manera de resolver esta disyuntiva.

30.5.2 El esquema IDA

El esquema implícito de dirección alternante, o esquema IDA, proporciona un medio para resolver ecuaciones parabólicas en dos dimensiones espaciales usando matrices tridiagonales. Para ello, cada incremento de tiempo se ejecuta en dos pasos (figura 30.10). En el primero, la ecuación (30.18) se aproxima mediante

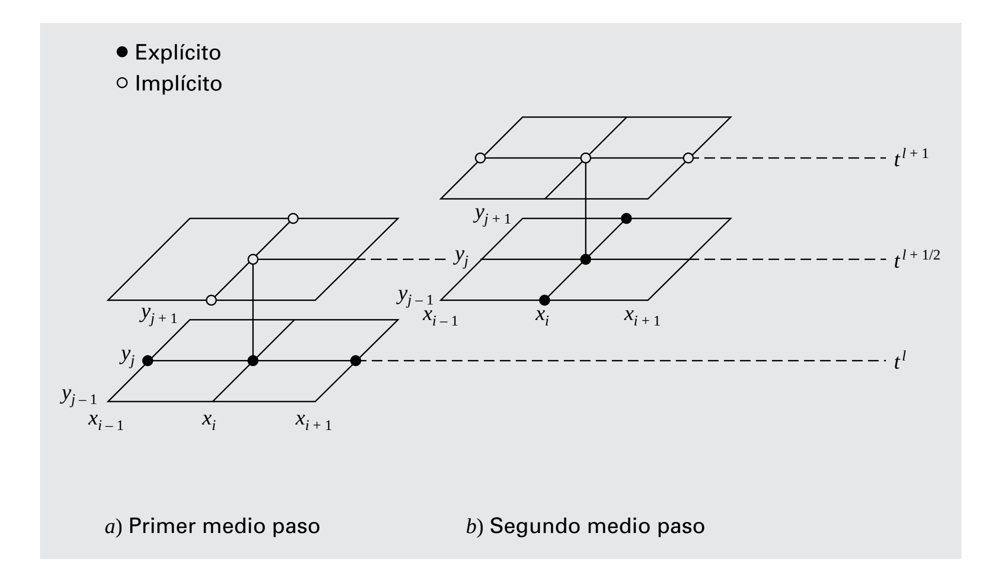


Figura 30.10
Los dos medios pasos usados en la implementación del esquema implícito de dirección alternante para resolver ecuaciones parabólicas en dos dimensiones espaciales.

$$\frac{T_{i,j}^{l+1/2}-T_{i,j}^{l}}{\Delta t/2}=k\left[\frac{T_{i+1,j}^{l}-2T_{i,j}^{l}+T_{i-1,j}^{l}}{(\Delta x)^2}+\frac{T_{i,j+1}^{l+1/2}-2T_{i,j}^{l+1/2}+T_{i,j-1}^{l+1/2}}{(\Delta y)^2}\right] \tag{30.19}$$

Así, la aproximación de $\partial^2T/\partial x^2$ se escribe explícitamente (es decir, en el punto base t^l, donde se conocen los valores de la temperatura). En consecuencia, sólo se desconocen tres términos de la temperatura en la aproximación de $\partial^2T/\partial y^2$. En el caso de una malla cuadrada ($\Delta y = \Delta x$), esta ecuación se expresa como

$$-\lambda T_{i,j-1}^{l+1/2}+2(1+\lambda)T_{i,j}^{l+1/2}-\lambda T_{i,j+1}^{l+1/2}=\lambda T_{i-1,j}^{l}+2(1-\lambda)T_{i,j}^{l}+\lambda T_{i+1,j}^{l} \tag{30.20}$$

que, cuando se escribe para el sistema, da como resultado un sistema tridiagonal de ecuaciones simultáneas.

En el segundo paso, que va desde $t^{l+1/2}$ hasta t^{l+1}, la ecuación (30.18) se aproxima por

$$\frac{T_{i,j}^{l+1}-T_{i,j}^{l+1/2}}{\Delta t/2}=k\left[\frac{T_{i+1,j}^{l+1}-2T_{i,j}^{l+1}+T_{i-1,j}^{l+1}}{(\Delta x)^2}+\frac{T_{i,j+1}^{l+1/2}-2T_{i,j}^{l+1/2}+T_{i,j-1}^{l+1/2}}{(\Delta y)^2}\right] \tag{30.21}$$

A diferencia de la ecuación (30.19), la aproximación de $\partial^2T/\partial x^2$ es ahora implícita. Así, el sesgo introducido por la ecuación (30.19) se corregirá parcialmente. Para una malla cuadrada, la ecuación (30.21) se escribe como

$$-\lambda T_{i-1,j}^{l+1}+2(1+\lambda)T_{i,j}^{l+1}-\lambda T_{i+1,j}^{l+1}=\lambda T_{i,j-1}^{l+1/2}+2(1-\lambda)T_{i,j}^{l+1/2}+\lambda T_{i,j+1}^{l+1/2} \tag{30.22}$$

De nuevo, cuando se escribe para una malla bidimensional, la ecuación da como resultado un sistema tridiagonal (figura 30.11). Como en el siguiente ejemplo, esto nos lleva a una solución numérica eficiente.

EJEMPLO 30.5 Método IDA

Planteamiento del problema Utilice el método IDA para encontrar la temperatura de la placa de los ejemplos 29.1 y 29.2. En $t = 0$, suponga que la temperatura de la placa es cero y que las temperaturas en la frontera se llevan instantáneamente a los niveles que se muestran en la figura 29.4. Emplee un tamaño de paso para el tiempo de 10 s. Recuerde del ejemplo 30.1 que el coeficiente de difusividad térmica para el aluminio es $k = 0.835$ cm²/s.

Solución Se empleó un valor de $\Delta x = 10$ cm para caracterizar la placa de 40 × 40 cm de los ejemplos 29.1 y 29.2. Por lo tanto, $\lambda = 0.835(10)/(10)^2 = 0.0835$. En el primer paso en $t = 5$ (figura 30.11*a*), la ecuación (30.20) se aplica a los nodos (1, 1), (1, 2) y (1, 3), que conducen a las siguientes ecuaciones tridiagonales:

$$\begin{bmatrix} 2.167 & -0.0835 & \\ -0.0835 & 2.167 & -0.0835 \\ & -0.0835 & 2.167 \end{bmatrix}\begin{Bmatrix} T_{1,1} \\ T_{1,2} \\ T_{1,3} \end{Bmatrix}=\begin{Bmatrix} 6.2625 \\ 6.2625 \\ 14.6125 \end{Bmatrix}$$

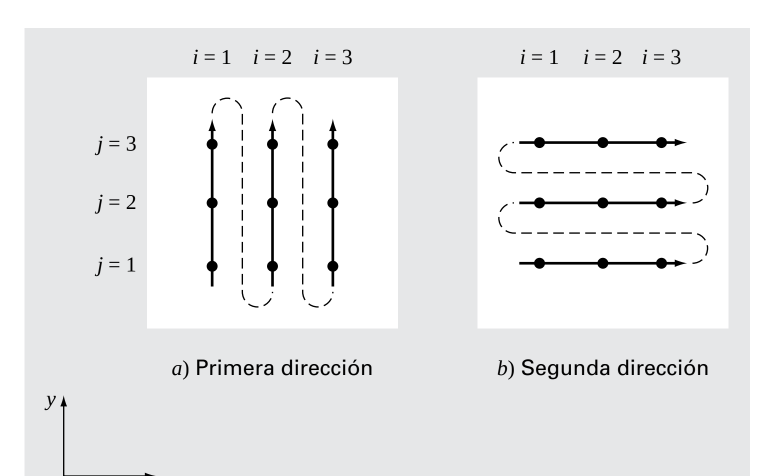
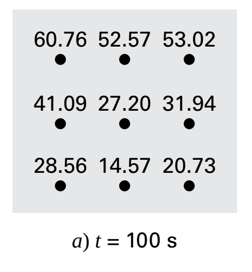
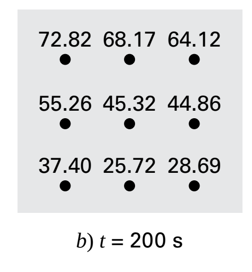


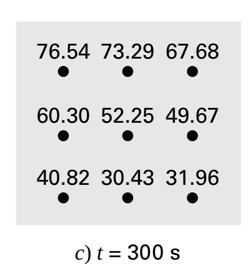
Figura 30.11
El método IDA da como resultado ecuaciones tridiagonales solamente si se aplica a lo largo de la dimensión que es implícita. En el primer paso *a*), se aplica a lo largo de la dimensión *y*; en el segundo paso *b*), a lo largo de la dimensión *x*. Estas "direcciones alternantes" son la razón del nombre del método.



a) t = 100 s



b) t = 200 s



c) t = 300 s

Figura 30.12
Solución para la placa calentada del ejemplo 30.5 en *a*) t = 100 s, *b*) t = 200 s y *c*) t = 300 s.

de las cuales se obtiene

$$T_{1,1} = 3.01597 \qquad T_{1,2} = 3.2708 \qquad T_{1,3} = 6.8692$$

De manera similar, se pueden desarrollar ecuaciones tridiagonales para encontrar

$$T_{2,1} = 0.1274 \qquad T_{2,2} = 0.2900 \qquad T_{2,3} = 4.1291$$

y

$$T_{3,1} = 2.0181 \qquad T_{3,2} = 2.2477 \qquad T_{3,3} = 6.0256$$

En el segundo paso en t = 10 (figura 30.11*b*), la ecuación (30.22) se aplica a los nodos (1, 1), (2, 1) y (3, 1) para llegar a

$$\begin{bmatrix} 2.167 & -0.0835 & \\ -0.0835 & 2.167 & -0.0835 \\ & -0.0835 & 2.167 \end{bmatrix} \begin{Bmatrix} T_{1,1} \\ T_{2,1} \\ T_{3,1} \end{Bmatrix} = \begin{Bmatrix} 12.0639 \\ 0.2577 \\ 8.0619 \end{Bmatrix}$$

para obtener

$$T_{1,1} = 5.5855 \qquad T_{2,1} = 0.4782 \qquad T_{3,1} = 3.7388$$

Ecuaciones tridiagonales para los otros renglones se desarrollan para llegar al siguiente resultado:

$$T_{1,2} = 6.1683 \qquad T_{2,2} = 0.8238 \qquad T_{3,2} = 4.2359$$

y

$$T_{1,3} = 13.1120 \qquad T_{2,3} = 8.3207 \qquad T_{3,3} = 11.3606$$

El cálculo puede repetirse, y los resultados para t = 100, 200 y 300 s se ilustran en las figuras 30.12*a* a 30.12*c*. Como se esperaba, la temperatura de la placa aumenta. Después de un espacio de tiempo suficiente, la temperatura se aproximará a la distribución en estado estacionario de la figura 29.5.

El método IDA es una de las técnicas de un grupo conocido como *métodos de división*. Algunos de éstos representan mejoras que evitan las desventajas del IDA. En muchas referencias se encuentran análisis de otros métodos de división, así como mayor información sobre el IDA (Ferziger, 1981; Lapidus y Pinder, 1981).

PROBLEMAS

30.1 Repita el ejemplo 30.1, pero utilice el método del punto medio para generar su solución.

30.2 Repita el ejemplo 30.1, pero para el caso en que la barra está inicialmente a 50°C y la derivada en x = 0 es igual a 1 y en x = 10 es igual a 0. Interprete sus resultados.

30.3 *a*) Repita el ejemplo 30.1, pero para un tiempo de paso Δt = 0.05 s. Compare los resultados con t = 0.2. *b*) Además, realice el mismo cálculo con el método de Heun (sin iteración del corrector) con un tamaño de paso mucho más pequeño, de Δt = 0.001 s. Suponga que los resultados del inciso *b*) son una aproximación válida de la

solución verdadera, y determine los errores relativos porcentuales para los resultados obtenidos en el ejemplo 30.1, así como para el inciso *a*).

30.4 Repita el ejemplo 30.2, pero para el caso en que la derivada en $x = 10$ es igual a cero.

30.5 Repita el ejemplo 30.3, pero para $\Delta x = 1$ cm.

30.6 Repita el ejemplo 30.5, pero para la placa descrita en el problema 29.2.

30.7 La ecuación de advección-difusión se utiliza para calcular la distribución de la concentración que hay en el lado largo de un reactor químico rectangular (véase la sección 32.1),

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} - U\frac{\partial c}{\partial x} - kc$$

donde c = concentración (mg/m^3), t = tiempo (min), D = coeficiente de difusión (m^2/min), x = distancia a lo largo del eje longitudinal del tanque (m), donde $x = 0$ en la entrada del tanque, U = velocidad en la dirección x (m/min) y k = tasa de reacción (min^{-1}) con la que el producto químico se convierte en otro. Desarrolle un esquema explícito para resolver esta ecuación en forma numérica. Pruébela para $k = 0.15$, $D = 100$ y $U = 1$, para un tanque con una longitud de 10 m. Use $\Delta x = 1$ m, y un tamaño de paso $\Delta t = 0.005$. Suponga que la concentración del flujo de entrada es de 100 y la concentración inicial en el tanque es de cero. Realice la simulación de $t = 0$ a 100 y grafique las concentraciones finales resultantes *versus* x.

30.8 Desarrolle un programa de cómputo amigable con el usuario para el método explícito simple de la sección 30.2. Pruébelo con la duplicación del ejemplo 30.1.

30.9 Modifique el programa del problema 30.8 de modo que emplee ya sea las condiciones de frontera de Dirichlet o derivadas. Pruébelo con la solución del problema 30.2.

30.10 Haga un programa de computadora amigable con el usuario para implantar el esquema implícito simple de la sección 30.3. Pruébelo con la duplicación del ejemplo 30.2.

30.11 Elabore un programa de computadora amigable con el usuario para implantar el método de Crank-Nicolson de la sección 30.4. Pruébelo con la duplicación del ejemplo 30.3.

30.12 Desarrolle un programa amigable con el usuario para el método IDA descrito en la sección 30.5. Pruébelo con la duplicación del ejemplo 30.5.

30.13 La forma no dimensional para la conducción de calor transitiva en una barra aislada (ecuación 30.1) se escribe como

$$\frac{\partial^2 u}{\partial \bar{x}^2} = \frac{\partial u}{\partial \bar{t}}$$

donde el espacio, tiempo y temperatura no dimensionales, se definen como

$$\bar{x} = \frac{x}{L} \qquad \bar{t} = \frac{T}{(\rho C L^2/k)} \qquad u = \frac{T - T_o}{T_L - T_o}$$

donde L = longitud de la barra, k = conductividad térmica del material de la barra, ρ = densidad, C = calor específico, T_0 = temperatura en $x = 0$ y T_L = temperatura en $x = L$. Esto opera para las siguientes condiciones iniciales y de frontera:

Condiciones de frontera	$u(0, \bar{t}) = 0$	$u(1, \bar{t}) = 1$
Condiciones iniciales	$u(\bar{x}, 0) = 0$	$0 \le \bar{x} \le 1$

Resuelva esta ecuación no dimensional para la distribución de la temperatura con los métodos de diferencias finitas y de Crank-Nicolson con una formulación exacta de segundo orden, para integrar en el tiempo. Escriba un programa de cómputo para obtener la solución. Incremente el valor de $\Delta\bar{t}$ en 10% para cada paso de tiempo para obtener con más rapidez la solución de estado estable, y seleccione valores de $\Delta\bar{x}$ y $\Delta\bar{t}$ para una exactitud buena. Grafique la temperatura no dimensional *versus* la longitud no dimensional para distintos valores de tiempos no dimensionales.

30.14 El problema del flujo de calor radial transitivo en una barra circular en forma no dimensional, está descrita por

$$\frac{\partial^2 u}{\partial \bar{r}^2} + \frac{1}{\bar{r}}\frac{\partial u}{\partial \bar{r}} = \frac{\partial u}{\partial \bar{t}}$$

Condiciones de frontera	$u(1, \bar{t}) = 1$	$\frac{\partial u}{\partial \bar{r}}(0, \bar{t}) = 0$
Condiciones iniciales	$u(\bar{x}, 0) = 0$	$0 \le \bar{x} \le 1$

Resuelva la ecuación de conducción de calor radial transitiva no dimensional en una barra circular para la distribución de temperatura en distintos tiempos conforme la temperatura de la barra se aproxima al estado estable. Utilice análogos de diferencias finitas exactas de segundo orden para las derivadas, con una formulación de Crank-Nicolson. Escriba un programa de computadora para la solución. Seleccione valores de $\Delta\bar{r}$ y $\Delta\bar{t}$ para una exactitud buena. Grafique la temperatura u *versus* el radio $\bar{r}$ para distintos tiempos $\bar{t}$.

30.15 Resuelva la siguiente EDP:

$$\frac{\partial^2 u}{\partial x^2} + b\frac{\partial u}{\partial x} = \frac{\partial u}{\partial t}$$

Condiciones de frontera	$u(0, t) = 0$	$u(1, t) = 1$
Condiciones iniciales	$u(x, 0) = 0$	$0 \le x \le 1$

Utilice análogos de diferencias finitas exactas de segundo orden para las derivadas, con una formulación de Crank-Nicolson, a fin de integrar en el tiempo. Escriba un programa de cómputo para obtener la solución. Incremente el valor de Δt en 10% para cada paso de tiempo a fin de obtener más rápido la solución de estado estable, y seleccione valores de Δx y Δt para una buena exactitud. Grafique u *versus* x para valores distintos de t. Resuelva para los valores de b = 4, 2, 0, –2, –4.

30.16 Determine las temperaturas a lo largo de una barra horizontal de 1 m, descritas por la ecuación de conducción del calor [ecuación (30.1)]. Suponga que la frontera derecha está aislada y que la izquierda ($x = 0$) está representada por

$$-k'\left.\frac{\partial T}{\partial x}\right|_{x=0} = h(T_a - T_0)$$

donde k' = coeficiente de conductividad térmica (W/m · °C), h = coeficiente de transferencia de calor convectivo (W/m^2 · °C), T_a = temperatura ambiente (°C) y T_0 = temperatura de la barra en $x = 0$ (°C). Resuelva cuál sería la temperatura como función del tiempo con el uso de un paso espacial de $\Delta x = 1$ cm y los siguientes valores de parámetros: $k = 2 \times 10^{-5}$ m^2/s, $k' = 10$ W/m · °C, $h = 25$ W/m^2 · °C, y $T_a = 50$ °C. Suponga que la temperatura inicial de la barra es cero.

CAPÍTULO 31

Método del elemento finito

Hasta aquí hemos empleado métodos por *diferencias finitas* para resolver ecuaciones diferenciales parciales. En estos métodos, el dominio de la solución se divide en una malla con puntos discretos o nodos (figura 31.1*b*). Entonces, se aplica la EDP en cada nodo, donde las derivadas parciales se reemplazan por diferencias finitas divididas. Aunque tal aproximación por puntos es conceptualmente fácil de entender, tiene varias desventajas. En particular, es difícil de aplicar a sistemas con una geometría irregular, con condiciones de frontera no usuales o de composición heterogénea.

El método del *elemento finito* ofrece una alternativa que es más adecuada para tales sistemas. A diferencia de las técnicas por diferencias finitas, la técnica del elemento finito divide el dominio de la solución en regiones con formas sencillas o "elementos" (figura 31.1*c*). Se puede desarrollar una solución aproximada de la EDP para cada uno de estos elementos. La solución total se genera uniendo, o "ensamblando", las soluciones individuales, teniendo cuidado de asegurar la continuidad de las fronteras entre los elementos. De este modo, la EDP se satisface *por secciones*.

Como se observa en la figura 31.1*c*, el uso de elementos, en lugar de una malla rectangular, proporciona una mejor aproximación para sistemas con forma irregular. Además, se pueden generar continuamente valores de las incógnitas a través de todo el dominio de la solución en lugar de puntos aislados.

Debido a que una descripción completa va más allá del alcance de este libro, el presente capítulo ofrece sólo una introducción general al método del elemento finito. Nuestro objetivo principal es familiarizar al lector con esta técnica y darle a conocer sus capacidades. Por lo tanto, la siguiente sección ofrece una visión general de los pasos para la solución de un problema, usando el elemento finito. Después se analizará un ejemplo sencillo: una barra caliente unidimensional en estado estacionario. Aunque este ejemplo no usa EDP, nos permite desarrollar y demostrar los principales

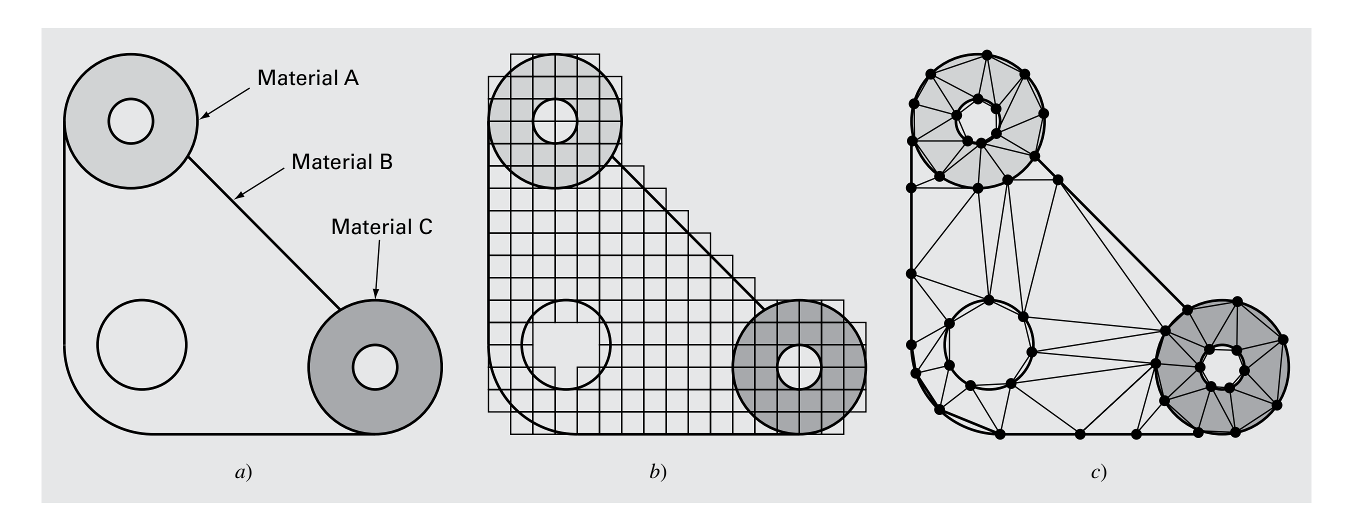


Figura 31.1
a) Un empaque con geometría irregular y composición no homogénea. *b*) Un sistema así es muy difícil de modelar con la técnica por diferencias finitas. Esto se debe al hecho de que se necesitan aproximaciones complicadas en las fronteras del sistema y en las fronteras entre las regiones de diferente composición. *c*) Una discretización por elementos finitos es mucho más adecuada para tales sistemas.

aspectos de la técnica del elemento finito, evitando llegar a factores complicados. Después podemos analizar algunos problemas con el empleo del método del elemento finito para resolver EDP.

31.1 EL ENFOQUE GENERAL

Aunque las particularidades varían, la implementación del método del elemento finito usualmente sigue un procedimiento estándar paso a paso. A continuación se presenta un panorama general de cada uno de estos pasos. La aplicación de tales pasos a problemas de ingeniería se desarrollará en las siguientes secciones.

31.1.1 Discretización

Este paso consiste en dividir el dominio de la solución en elementos finitos. En la figura 31.2 se muestran ejemplos de los elementos empleados en una, dos y tres dimensiones. Los puntos de intersección de las líneas que forman los lados de los elementos se conocen como nodos, y los mismos lados se denominan *líneas* o *planos nodales*.

31.1.2 Ecuaciones de los elementos

El siguiente paso consiste en desarrollar ecuaciones para aproximar la solución de cada elemento y consta de dos pasos. Primero, se debe elegir una función apropiada con coeficientes desconocidos que aproximará la solución. Segundo, se evalúan los coeficientes de modo que la función aproxime la solución de manera óptima.

Elección de las funciones de aproximación Debido a que son fáciles de manipular matemáticamente, a menudo se utilizan polinomios para este propósito. En el caso unidimensional, la alternativa más sencilla es un polinomio de primer grado o línea recta.

$$u(x) = a_0 + a_1 x \tag{31.1}$$

donde $u(x)$ = la variable dependiente, a_0 y a_1 = constantes y x = la variable independiente. Esta función debe pasar a través de los valores de $u(x)$ en los puntos extremos del elemento en x_1 y x_2. Por lo tanto,

$$u_1 = a_0 + a_1 x_1$$
$$u_2 = a_0 + a_1 x_2$$

donde $u_1 = u(x_1)$ y $u_2 = u(x_2)$. De estas ecuaciones, usando la regla de Cramer, se obtiene

$$a_0 = \frac{u_1 x_2 - u_2 x_1}{x_2 - x_1} \qquad a_1 = \frac{u_2 - u_1}{x_2 - x_1}$$

Estos resultados se sustituyen en la ecuación (31.1) la cual, después de reagrupar términos, se escribe como

$$u = N_1 u_1 + N_2 u_2 \tag{31.2}$$

donde

$$N_1 = \frac{x_2 - x}{x_2 - x_1} \tag{31.3}$$

y

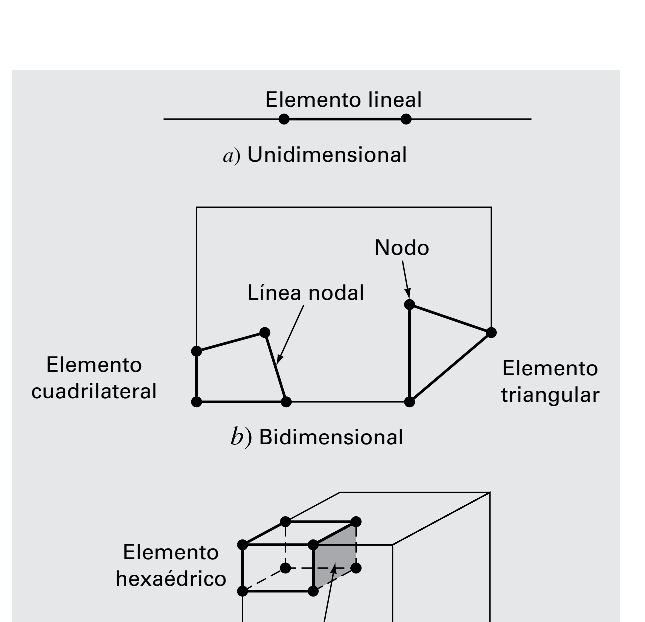


Figura 31.2
Ejemplos de los elementos empleados en *a*) una, *b*) dos y *c*) tres dimensiones.

$$N_2 = \frac{x - x_1}{x_2 - x_1} \tag{31.4}$$

La ecuación (31.2) se conoce como *una función de aproximación*, o *de forma*, y N_1 y N_2 se *denominan funciones de interpolación*. Una inspección cuidadosa revela que la ecuación (31.2) es, en realidad, el polinomio de interpolación de primer grado de Lagrange. Esta ecuación ofrece un medio para predecir valores intermedios (es decir, para interpolar) entre valores dados u_1 y u_2 en los nodos.

La figura 31.3 muestra la función de forma junto con las funciones de interpolación correspondientes. Observe que la suma de las funciones de interpolación es igual a uno.

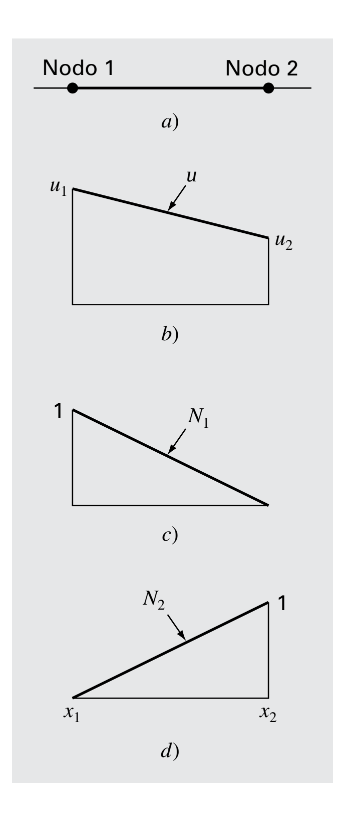


Figura 31.3
b) Una aproximación lineal o función de forma para *a*) un elemento lineal. Las funciones de interpolación correspondientes se muestran en *c*) y *d*).

Además, el hecho de que estemos tratando con ecuaciones lineales facilita las operaciones como la diferenciación y la integración. Tales manipulaciones serán importantes en secciones posteriores. La derivada de la ecuación (31.2) es

$$\frac{du}{dx} = \frac{dN_1}{dx}u_1 + \frac{dN_2}{dx}u_2 \tag{31.5}$$

De acuerdo con las ecuaciones (31.3) y (31.4), las derivadas de las N se calculan como sigue

$$\frac{dN_1}{dx} = -\frac{1}{x_2 - x_1} \qquad \frac{dN_2}{dx} = \frac{1}{x_2 - x_1} \tag{31.6}$$

y, por lo tanto, la derivada de u es

$$\frac{du}{dx} = \frac{1}{x_2 - x_1}(-u_1 + u_2) \tag{31.7}$$

En otras palabras, es una diferencia dividida que representa la pendiente de la línea recta que une los nodos.

La integral se expresa como

$$\int_{x_1}^{x_2} u\,dx = \int_{x_1}^{x_2} \left(N_1u_1 + N_2u_2\right)dx$$

Cada uno de los términos del lado derecho es simplemente la integral de un triángulo rectángulo con base $x_2 - x_1$ y altura u. Es decir,

$$\int_{x_1}^{x_2} Nu\,dx = \frac{1}{2}(x_2 - x_1)u$$

Así, la integral completa es

$$\int_{x_1}^{x_2} u\,dx = \frac{u_1 + u_2}{2}(x_2 - x_1) \tag{31.8}$$

En otras palabras, esto es simplemente la regla del trapecio.

Obtención de un ajuste óptimo de la función a la solución Una vez que se ha elegido la función de interpolación, se debe desarrollar la ecuación que rige el comportamiento del elemento. Esta ecuación representa un ajuste de la función a la solución de la ecuación diferencial de que se trate. Existen varios métodos para este propósito; entre los más comunes están el método directo, el método de los residuos ponderados y el método variacional. Los resultados de todos estos métodos son análogos al ajuste de curvas. Sin embargo, en lugar de ajustar funciones a datos, estos métodos especifican relaciones entre las incógnitas de la ecuación (31.2) que satisfacen de manera óptima la EDP.

Matemáticamente, las ecuaciones del elemento resultante a menudo consisten en un sistema de ecuaciones algebraicas lineales que puede expresarse en forma matricial,

$$[k]\{u\} = \{F\} \tag{31.9}$$

donde $[k]$ = *una propiedad del elemento o matriz de rigidez,* $\{u\}$ = vector columna de las incógnitas en los nodos y $\{F\}$ = vector columna determinado por el efecto de cualquier influencia externa aplicada a los nodos. Observe que, en algunos casos, las ecuaciones pueden ser no lineales. Sin embargo, en los ejemplos elementales descritos aquí, así como en muchos problemas prácticos, los sistemas son lineales.

31.1.3 Ensamble

Una vez obtenidas las ecuaciones de elementos individuales, éstas deben unirse o ensamblarse para caracterizar el comportamiento de todo el sistema. El proceso de ensamble está regido por el concepto de continuidad. Es decir, las soluciones de elementos contiguos se acoplan, de manera que los valores de las incógnitas (y algunas veces las derivadas) en sus nodos comunes sean equivalentes. Así, la solución total será continua.

Cuando finalmente todas las versiones individuales de la ecuación (31.9) están ensambladas, el sistema completo se expresa en forma matricial como

$$[K]\{u'\} = \{F'\} \tag{31.10}$$

donde $[K]$ = *la matriz de propiedades de ensamble* y $\{u'\}$ y $\{F'\}$ = vectores columna de las incógnitas y de las fuerzas externas, marcadas con apóstrofos para denotar que son ensamble de los vectores {u} y $\{F\}$ de los elementos individuales.

31.1.4 Condiciones de frontera

Antes de resolver la ecuación (31.10) debe modificarse para considerar las condiciones de frontera del sistema. Dichos ajustes dan como resultado

$$[\bar{k}]\{u'\} = \{\bar{F}'\} \tag{31.11}$$

donde la barra significa que las condiciones de frontera se han incorporado.

31.1.5 Solución

Las soluciones de la ecuación (31.11) se obtienen con las técnicas que se describieron en la parte tres, tal como la descomposición *LU.* En muchos casos, los elementos pueden configurarse de manera que en las ecuaciones resultantes su matriz de coeficientes sea de banda. Así, es posible utilizar los esquemas de solución altamente eficientes para estos sistemas.

31.1.6 Procesamiento posterior

Una vez obtenida la solución, ésta se despliega en forma tabular o de manera gráfica. Además, pueden determinarse las variables secundarias y también mostrarse.

Aunque los pasos anteriores son muy generales, son comunes a la mayoría de las implementaciones del método del elemento finito. En la siguiente sección ilustraremos cómo se aplican para obtener resultados numéricos de un sistema físico simple (una barra caliente).

31.2 APLICACIÓN DEL ELEMENTO FINITO EN UNA DIMENSIÓN

En la figura 31.4 se muestra un sistema que puede modelarse mediante la forma unidimensional de la ecuación de Poisson

$$\frac{d^2T}{dx^2} = -f(x) \tag{31.12}$$

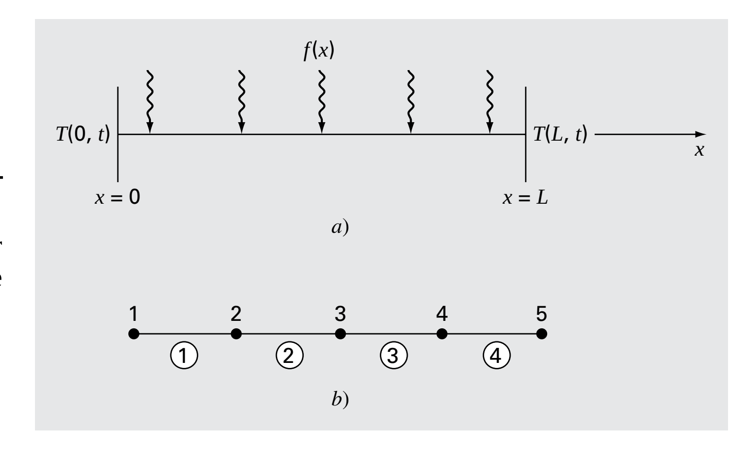


Figura 31.4
a) Barra larga y delgada sujeta a condiciones de frontera fijas y una fuente de calor continua a lo largo de su eje. *b*) Representación del elemento finito que consta de cuatro elementos de igual longitud y cinco nodos.

donde $f(x)$ = una función que define una fuente de calor a lo largo de la barra, y donde los extremos de la barra se mantienen a temperaturas fijas,

$$T(0, t) = T_1$$

y

$$T(L, t) = T_2$$

Observe que ésta no es una ecuación diferencial parcial, sino una EDO con valor en la frontera. Se usa este modelo sencillo porque nos permitirá introducir el método del elemento finito sin algunas de las complicaciones de una EDP bidimensional, por ejemplo.

EJEMPLO 31.1 Solución analítica para una barra caliente

Planteamiento del problema Resuelva la ecuación (31.12) para una barra de 10 cm con las siguientes condiciones de frontera, $T(0, t) = 40$ y $T(10, t) = 200$ y una fuente de calor uniforme de $f(x) = 10$.

Solución La ecuación a resolver es

$$\frac{d^2T}{dx^2} = -10$$

Suponga una solución de la forma

$$T = ax^2 + bx + c$$

la cual se deriva dos veces para obtener $T'' = 2a$. Sustituyendo este resultado en la ecuación diferencial da $a = -5$. Las condiciones de frontera se utilizan para evaluar los coeficientes restantes. Para la primera condición en $x = 0$,

$$40 = -5(0)^2 + b(0) + c$$

o $c = 40$. De manera similar, para la segunda condición,

$$200 = -5(10)^2 + b(10) + 40$$

de donde se obtiene $b = 66$. Por lo tanto, la solución final es

$$T = -5x^2 + 66x + 40$$

Los resultados se grafican en la figura 31.5.

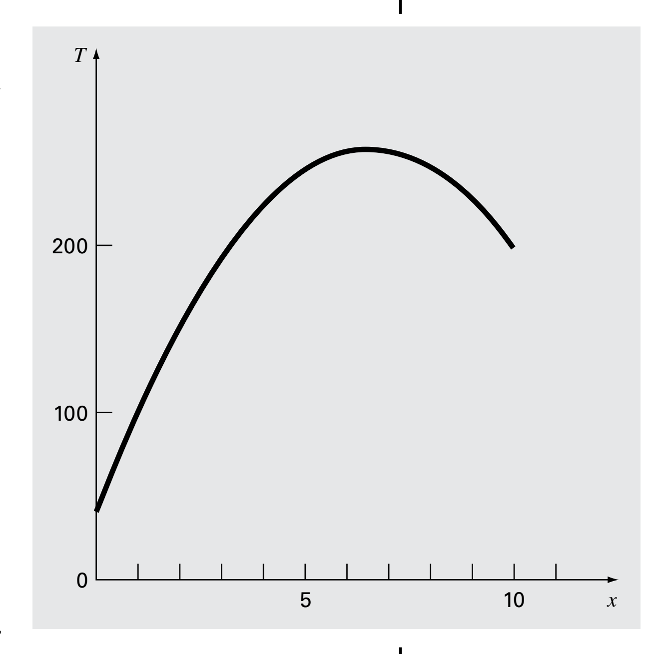


Figura 31.5
Distribución de temperatura a lo largo de una placa caliente sujeta a una fuente de calor uniforme y mantenida a temperaturas fijas en los extremos.

31.2.1 Discretización

Una configuración sencilla para modelar el sistema consiste en una serie de elementos de igual longitud (figura 31.4*b*). Así, el sistema se trata como cuatro elementos de igual longitud y cinco nodos.

31.2.2 Ecuaciones de los elementos

En la figura 31.6*a* se muestra un elemento individual. La distribución de temperatura para el elemento se representa por la función de aproximación

$$\tilde{T} = N_1T_1 + N_2T_2 \tag{31.13}$$

donde N_1 y N_2 = funciones de interpolación lineales especificadas por las ecuaciones (31.3) y (31.4), respectivamente. De esta manera, como se ilustra en la figura 31.6*b*, la función de aproximación corresponde a una interpolación lineal entre las dos temperaturas nodales.

Como se presentó en la sección 31.1, existen diferentes métodos para desarrollar la ecuación del elemento. En esta sección empleamos dos de ellos. Primero, se usará un *método directo* para el caso sencillo donde $f(x) = 0$. Posteriormente, debido a su aplicabilidad general en ingeniería, dedicamos la mayor parte de la sección al *método de los residuos ponderados*.

El método directo En el caso donde $f(x) = 0$, se utiliza un método directo para generar las ecuaciones de los elementos. La relación entre el flujo de calor y el gradiente de temperatura puede representarse mediante la ley de Fourier:

$$q = -k'\frac{dT}{dx}$$

donde q = flujo [cal/(cm^2 · s)] y k' = coeficiente de conductividad térmica [cal/(s · cm · °C)]. Si se utiliza una función de aproximación lineal para caracterizar la temperatura del elemento, el flujo de calor hacia el elemento a través del nodo 1 se representa por

$$q_1 = k'\frac{T_1 - T_2}{x_2 - x_1}$$

donde q_1 es el flujo de calor en el nodo 1. De manera similar, para el nodo 2,

$$q_2 = k'\frac{T_2 - T_1}{x_2 - x_1}$$

Estas dos ecuaciones expresan la relación de la distribución de la temperatura interna de los elementos (determinada por las temperaturas nodales) con el flujo de calor en sus extremos. En consecuencia representan nuestras ecuaciones deseadas de los elementos. Se simplifican aún más reconociendo que la ley de Fourier se puede utilizar para expresar los flujos de los extremos en términos de los gradientes de temperatura en las fronteras. Es decir,

$$q_1 = -k'\frac{dT(x_1)}{dx} \qquad q_2 = k'\frac{dT(x_2)}{dx}$$

que se sustituyen en las ecuaciones de los elementos para dar

$$\frac{1}{x_2 - x_1}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} T_1 \\ T_2 \end{Bmatrix} = \begin{Bmatrix} -\dfrac{dT(x_1)}{dx} \\ \dfrac{dT(x_2)}{dx} \end{Bmatrix} \tag{31.14}$$

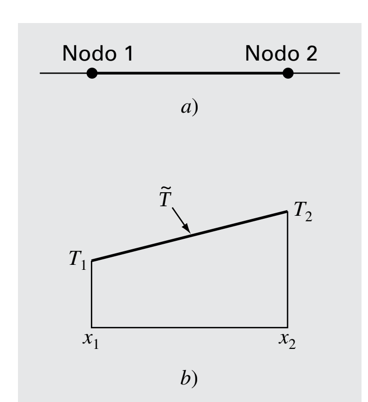


Figura 31.6
a) Un elemento individual. *b*) Función de aproximación usada para caracterizar la distribución de temperatura a lo largo del elemento.

Observe que la ecuación (31.14) se presentó en el formato de la ecuación (31.9). Así, logramos generar una ecuación matricial que describe el comportamiento de un elemento típico de nuestro sistema.

El método directo resulta muy intuitivo. Se utiliza en áreas como la mecánica, para resolver problemas importantes. Sin embargo, en otros contextos a menudo es difícil, si no es que imposible, obtener directamente las ecuaciones del elemento finito. En consecuencia, como se describe a continuación, se cuenta con técnicas matemáticas más generales.

El método de los residuos ponderados La ecuación diferencial (31.12) se reexpresa como

$$0 = \frac{d^2T}{dx^2} + f(x)$$

La solución aproximada [ecuación (31.13)] se sustituye en esta ecuación. Como la ecuación (31.13) no es la solución exacta, el lado izquierdo de la ecuación resultante no será cero, sino que será igual a un residuo,

$$R = \frac{d^2\tilde{T}}{dx^2} + f(x) \tag{31.15}$$

El *método de los residuos ponderados* (MRP) consiste en encontrar un mínimo para el residuo, de acuerdo con la fórmula general

$$\int_D RW_i \, dD = 0 \qquad i = 1, 2, \ldots, m \tag{31.16}$$

donde D = dominio de la solución y W_i = funciones de ponderación linealmente independientes.

Aquí, se tienen múltiples opciones para las funciones de ponderación (cuadro 31.1). El procedimiento más común para el método del elemento finito consiste en emplear las funciones de interpolación N_i como las funciones de ponderación. Cuando éstas se sustituyen en la ecuación (31.16), el resultado se conoce como el método de Galerkin,

Cuadro 31.1 Esquemas de residuos alternativos para el MRP

Se puede elegir entre varias funciones de ponderación para la ecuación (31.16). Cada una representa un procedimiento alternativo para el MRP.

En el *método de la colocación*, elegimos tantas posiciones como coeficientes desconocidos existan. Después, se ajustan los coeficientes hasta que los residuos desaparezcan en cada una de estas posiciones. Por lo tanto, la función de aproximación dará resultados perfectos en las posiciones elegidas, pero en las posiciones restantes tendremos un residuo diferente de cero. Así, este método es parecido a los de interpolación del capítulo 18. Observe que la colocación corresponde a usar la función de ponderación

$$W = \delta\,(x - x_i) \qquad \text{para } i = 1, 2, ..., n$$

donde n = el número de coeficientes desconocidos y $\delta\,(x - x_i)$ = la *función delta de Dirac*, que es igual a cero en todas partes excepto en $x = x_i$, donde es igual a 1.

En el *método del subdominio*, el intervalo se divide en tantos segmentos, o "subdominios", como coeficientes desconocidos existan. Después, se ajustan los coeficientes hasta que el valor promedio del residuo sea cero en cada subdominio. Así, en cada subdominio, la función de ponderación será igual a 1, y la ecuación (31.16) se convierte en

$$\int_{x_{i-1}}^{x_i} R \, dx = 0 \qquad \text{para } i = 1, 2, \ldots, n$$

donde x_{i-1} y x_i son las fronteras del subdominio.

En el caso de *mínimos cuadrados*, los coeficientes se ajustan hasta minimizar la integral del cuadrado del residuo. De manera que las funciones de ponderación son

$$W_i = \frac{\partial R}{\partial a_i}$$

al sustituir las W_i en la ecuación (31.16), se obtiene

$$\int_D R \frac{\partial R}{\partial a_i} \, dD = 0 \qquad i = 1, 2, \ldots, n$$

o

$$\frac{\partial}{\partial a_i} \int_D R^2 \, dD = 0 \qquad i = 1, 2, \ldots, n$$

La comparación de esta formulación con la del capítulo 17 muestra que ésta es la forma continua de la regresión.

El método de Galerkin emplea las funciones de interpolación N_i como funciones de ponderación. Recuerde que estas funciones siempre suman 1 en cualquier posición en un elemento. En muchos problemas, el método de Galerkin da los mismos resultados que los que se obtienen con los métodos variacionales. En consecuencia, ésta es la versión del MRP que se emplea con más frecuencia en el análisis del elemento finito.

$$\int_D RN_i \, dD = 0 \qquad i = 1, 2, \ldots, m$$

En nuestra barra unidimensional, la ecuación (31.15) se sustituye en esta formulación para dar

$$\int_{x_1}^{x_2} \left[\frac{d^2\tilde{T}}{dx^2} + f(x) \right] N_i \, dx \quad i = 1, 2$$

que se pueden reexpresar como sigue:

$$\int_{x_1}^{x_2} \frac{d^2\tilde{T}}{dx^2} N_i(x) \, dx = -\int_{x_1}^{x_2} f(x) N_i(x) \, dx \qquad i = 1, 2 \tag{31.17}$$

Ahora, se aplicarán varias manipulaciones matemáticas para simplificar y evaluar la ecuación (31.17). Una de las más importantes es la simplificación del lado izquierdo usando la integración por partes. Del cálculo, recuerde que esta operación se expresa como

$$\int_a^b u \, dv = uv\big|_a^b - \int_a^b v \, du$$

Si u y v se eligen adecuadamente, la nueva integral en el lado derecho será más fácil de evaluar que la integral original del lado izquierdo. Esto se puede hacer para el término del lado izquierdo de la ecuación (31.17), escogiendo $N_i(x)$ como u, y $(d^2T/dx^2)\,dx$ como dv, se obtiene

$$\int_{x_1}^{x_2} N_i(x) \frac{d^2\tilde{T}}{dx^2} dx = N_i(x) \frac{d\tilde{T}}{dx}\bigg|_{x_1}^{x_2} - \int_{x_1}^{x_2} \frac{d\tilde{T}}{dx} \frac{dN_i}{dx} dx \qquad i = 1, 2 \tag{31.18}$$

Así, hemos dado el importante paso de bajar el orden superior en la formulación: de una segunda a una primera derivada.

A continuación, se evalúa cada uno de los términos que hemos creado en la ecuación (31.18). Para $i = 1$, el primer término del lado derecho de la ecuación (31.18) se evalúa como sigue

$$N_1(x) \frac{d\tilde{T}}{dx}\bigg|_{x_1}^{x_2} = N_1(x_2) \frac{d\tilde{T}(x_2)}{dx} - N_1(x_1) \frac{d\tilde{T}(x_1)}{dx}$$

Sin embargo, de la figura 31.3 recuerde que $N_1(x_2) = 0$ y $N_1(x_1) = 1$ y, por lo tanto,

$$N_1(x) \frac{d\tilde{T}}{dx}\bigg|_{x_1}^{x_2} = -\frac{d\tilde{T}(x_1)}{dx} \tag{31.19}$$

De manera similar, para $i = 2$,

$$N_2(x) \frac{d\tilde{T}}{dx}\bigg|_{x_1}^{x_2} = \frac{d\tilde{T}(x_2)}{dx} \tag{31.20}$$

Así, el primer término en el lado derecho de la ecuación (31.18) representa las condiciones de frontera naturales en los extremos de los elementos.

Ahora, antes de continuar, reagrupemos sustituyendo en la ecuación original los términos correspondientes por nuestros resultados. Empleamos las ecuaciones (31.18) a (31.20) para hacer las sustituciones correspondientes en la ecuación (31.17); para $i = 1$,

$$\int_{x_1}^{x_2} \frac{d\tilde{T}}{dx} \frac{dN_1}{dx} dx = -\frac{d\tilde{T}(x_1)}{dx} + \int_{x_1}^{x_2} f(x) N_1(x) \, dx \tag{31.21}$$

y para $i = 2$,

$$\int_{x_1}^{x_2} \frac{d\tilde{T}}{dx}\frac{dN_2}{dx}\,dx = \frac{d\tilde{T}(x_2)}{dx} + \int_{x_1}^{x_2} f(x)N_2(x)\,dx \tag{31.22}$$

Observe que la integración por partes nos llevó a dos importantes resultados. Primero, ha incorporado las condiciones de frontera directamente dentro de las ecuaciones del elemento. Segundo, ha bajado la evaluación de orden superior, de una segunda a una primera derivada. Este último resultado tiene como consecuencia significativa que las funciones de aproximación necesitan preservar continuidad de valor, pero no pendiente en los nodos.

Observe también que ahora podemos comenzar a darles significado físico a cada uno de los términos que obtuvimos. En el lado derecho de cada ecuación, el primer término representa una de las condiciones de frontera del elemento; y el segundo es el efecto de la función de fuerza del sistema, en este caso, la fuente de calor $f(x)$. Como ahora será evidente, el lado izquierdo representa los mecanismos internos que rigen la distribución de la temperatura del elemento. Es decir, en términos del método del elemento finito, el lado izquierdo será la matriz de propiedad del elemento.

Para ver esto nos concentramos en los términos del lado izquierdo. Para $i = 1$, el término es

$$\int_{x_1}^{x_2} \frac{d\tilde{T}}{dx}\frac{dN_1}{dx}\,dx \tag{31.23}$$

Recordemos de la sección 31.1.2 que la naturaleza lineal de la función hace que la diferenciación y la integración sean sencillas. Si empleamos las ecuaciones (31.6) y (31.7) para hacer las sustituciones correspondientes en la ecuación (31.23), obtenemos

$$\int_{x_1}^{x_2} \frac{T_1 - T_2}{(x_2 - x_1)^2}\,dx = \frac{1}{x_2 - x_1}(T_1 - T_2) \tag{31.24}$$

De manera similar para $i = 2$ [ecuación (31.22)],

$$\int_{x_1}^{x_2} \frac{-T_1 + T_2}{(x_2 - x_1)^2}\,dx = \frac{1}{x_2 - x_1}(-T_1 + T_2) \tag{31.25}$$

Una comparación con la ecuación (31.14) nos muestra que éstas son similares a las relaciones obtenidas con el método directo usando la ley de Fourier, lo cual se aclara más al expresar las ecuaciones (31.24) y (31.25) en forma matricial como sigue:

$$\frac{1}{x_2 - x_1}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} T_1 \\ T_2 \end{Bmatrix}$$

Si este resultado se sustituye en las ecuaciones (31.21) y (31.22), y después se expresa en forma matricial, obtenemos la versión final de las ecuaciones de los elementos.

$$\underbrace{\frac{1}{x_2 - x_1}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}}_{\substack{\text{Matriz de rigidez}\\ \text{del elemento}}}\{T\} = \underbrace{\begin{Bmatrix} -\dfrac{dT(x_1)}{dx} \\ \dfrac{dT(x_2)}{dx} \end{Bmatrix}}_{\substack{\text{Condición}\\ \text{de frontera}}} + \underbrace{\begin{Bmatrix} \displaystyle\int_{x_1}^{x_2} f(x)N_1(x)\,dx \\ \displaystyle\int_{x_1}^{x_2} f(x)N_2(x)\,dx \end{Bmatrix}}_{\text{Efectos externos}} \tag{31.26}$$

Observe que las ecuaciones del elemento pueden obtenerse no sólo mediante los métodos directo y de los residuos ponderados, sino también usando el cálculo de variaciones (por ejemplo, véase Allaire, 1985). En el caso presente, este método proporciona ecuaciones idénticas a las deducidas arriba.

EJEMPLO 31.2 Ecuación del elemento en una barra caliente

Planteamiento del problema Emplee la ecuación (31.26) para desarrollar las ecuaciones del elemento dada una barra de 10 cm, con condiciones de frontera de $T(0, t) = 40$ y $T(10, t) = 200$ y una fuente de calor uniforme con $f(x) = 10$. Utilice cuatro elementos del mismo tamaño con longitud = 2.5 cm.

Solución El término de la fuente de calor en el primer renglón de la ecuación (31.26) se evalúa sustituyendo la ecuación (31.3), e integrando para obtener

$$\int_0^{2.5} 10\frac{2.5-x}{2.5}\,dx = 12.5$$

De manera similar, la ecuación (31.4) se sustituye en el término de la fuente de calor del segundo renglón de la ecuación (31.26), el cual también se integra para obtener

$$\int_0^{2.5} 10\frac{x-0}{2.5}\,dx = 12.5$$

Estos resultados, junto con los valores de los otros parámetros, se emplean para sustituirse en la ecuación (31.26) y así obtener

$$0.4T_1 - 0.4T_2 = -\frac{dT}{dx}(x_1) + 12.5$$

y

$$-0.4T_1 + 0.4T_2 = \frac{dT}{dx}(x_2) + 12.5$$

31.2.3 Ensamble

Antes de que se ensamblen las ecuaciones del elemento, se debe establecer un esquema de numeración global que especifique la topología o el arreglo espacial del sistema. Como en la tabla 31.1, esto define la conectividad de los elementos en la malla. Debido a que este caso es unidimensional, el esquema de numeración parecerá tan predecible que resulta trivial. Sin embargo, en problemas para dos y tres dimensiones, tal esquema es el único medio para especificar qué nodos pertenecen a qué elementos.

Una vez que se ha especificado la topología, la ecuación del elemento (31.26) se puede escribir para otros elementos usando las coordenadas del sistema. Después, éstas se agregan (una por una) para ensamblar la matriz de todo el sistema (este proceso se continuará explorando en la sección 32.4). El proceso se ilustra en la figura 31.7.

Tabla 31.1 Topología del sistema para el esquema de segmentación del elemento finito de la figura 31.4*b*.

	Número de nodos	
Elemento	**Local**	**Global**
1	1	1
	2	2
2	1	2
	2	3
3	1	3
	2	4
4	1	4
	2	5

31.2.4 Condiciones de la frontera

Observe que conforme se ensamblan las ecuaciones, se cancelan las condiciones de frontera internas. Así, el resultado final de {F} en la figura 31.7*e* tiene condiciones de frontera sólo para el primero y el último nodos. Ya que T_1 y T_5 están dados, dichas condiciones de frontera naturales en los extremos de la barra, $dT(x_1)/dx$ y $dT(x_5)/dx$, representan incógnitas. Por lo tanto, las ecuaciones se reexpresan como sigue:

$$
a)\;\begin{bmatrix} 0.4 & -0.4 & 0 & 0 & 0 \\ -0.4 & 0.4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ 0 \\ 0 \\ 0 \end{Bmatrix} = \begin{Bmatrix} -dT(x_1)/dx + 12.5 \\ dT(x_2)/dx + 12.5 \\ 0 \\ 0 \\ 0 \end{Bmatrix}
$$

$$
b)\;\begin{bmatrix} 0.4 & -0.4 & & 0 & 0 & 0 \\ -0.4 & 0.4 & \boxed{+0.4} & \boxed{-0.4} & 0 & 0 \\ 0 & & \boxed{-0.4} & \boxed{0.4} & 0 & 0 \\ 0 & 0 & & 0 & 0 & 0 \\ 0 & 0 & & 0 & 0 & 0 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \\ 0 \\ 0 \end{Bmatrix} = \begin{Bmatrix} -dT(x_1)/dx + 12.5 \\ 12.5 + 12.5 \\ dT(x_3)/dx + 12.5 \\ 0 \\ 0 \end{Bmatrix}
$$

$$
c)\;\begin{bmatrix} 0.4 & -0.4 & 0 & & 0 & 0 \\ -0.4 & 0.8 & -0.4 & & 0 & 0 \\ 0 & -0.4 & 0.4 & \boxed{+0.4} & \boxed{-0.4} & 0 \\ 0 & 0 & & \boxed{-0.4} & \boxed{0.4} & 0 \\ 0 & 0 & 0 & & 0 & 0 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ 0 \end{Bmatrix} = \begin{Bmatrix} -dT(x_1)/dx + 12.5 \\ 25 \\ 12.5 + 12.5 \\ dT(x_4)/dx + 12.5 \\ 0 \end{Bmatrix}
$$

$$
d)\;\begin{bmatrix} 0.4 & -0.4 & 0 & 0 & & 0 \\ -0.4 & 0.8 & -0.4 & 0 & & 0 \\ 0 & -0.4 & 0.8 & -0.4 & & 0 \\ 0 & 0 & -0.4 & 0.4 & \boxed{+0.4} & \boxed{-0.4} \\ 0 & 0 & 0 & & \boxed{-0.4} & \boxed{0.4} \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \end{Bmatrix} = \begin{Bmatrix} -dT(x_1)/dx + 12.5 \\ 25 \\ 25 \\ 12.5 + 12.5 \\ dT(x_5)/dx + 12.5 \end{Bmatrix}
$$

$$
e)\;\begin{bmatrix} 0.4 & -0.4 & 0 & 0 & 0 \\ -0.4 & 0.8 & -0.4 & 0 & 0 \\ 0 & -0.4 & 0.8 & -0.4 & 0 \\ 0 & 0 & -0.4 & 0.8 & -0.4 \\ 0 & 0 & 0 & -0.4 & 0.4 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \end{Bmatrix} = \begin{Bmatrix} -dT(x_1)/dx + 12.5 \\ 25 \\ 25 \\ 25 \\ dT(x_5)/dx + 12.5 \end{Bmatrix}
$$

Figura 31.7
Ensamble de las ecuaciones de todo el sistema.

$$
\begin{array}{lllllcl}
\dfrac{dT}{dx}(x_1) & -0.4T_2 & & & & = & -3.5 \\
 & 0.8T_2 & -0.4T_3 & & & = & 41 \\
 & -0.4T_2 & +0.8T_3 & -0.4T_4 & & = & 25 \\
 & & -0.4T_3 & +0.8T_4 & & = & 105 \\
 & & & -0.4T_4 & -\dfrac{dT}{dx}(x_5) & = & -67.5
\end{array}
\tag{31.27}
$$

31.2.5 Solución

De la ecuación (31.27) se obtiene

$$
\frac{dT}{dx}(x_1) = 66 \qquad T_2 = 173.75 \qquad T_3 = 245
$$

$$
T_4 = 253.75 \qquad \frac{dT}{dx}(x_5) = -34
$$

31.2.6 Proceso posterior

Los resultados se representan gráficamente. En la figura 31.8 se muestran los resultados del método del elemento finito, junto con la solución exacta. Observe que el cálculo del elemento finito capta la

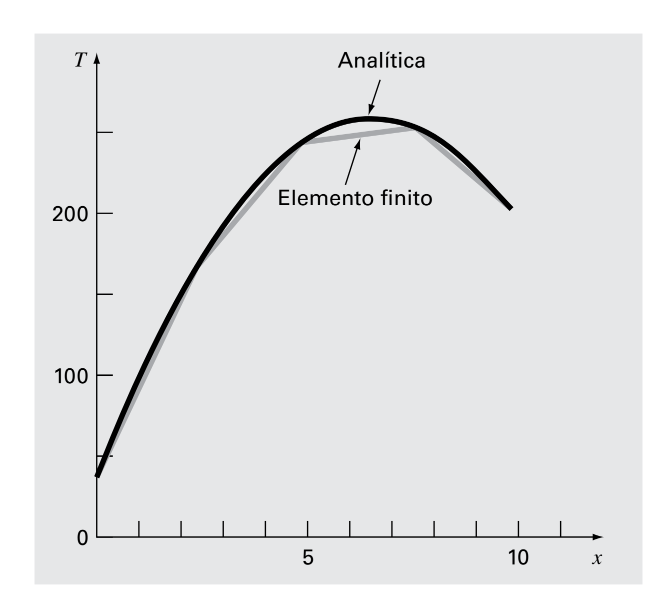


Figura 31.8
Resultados al aplicar el método del elemento finito a una barra caliente. También se muestra la solución exacta.

tendencia general de la solución exacta y, además, da una coincidencia exacta en los nodos. Sin embargo, existe una discrepancia en el interior de cada elemento debido a la naturaleza lineal de las funciones de forma.

31.3 PROBLEMAS BIDIMENSIONALES

Aunque la "contabilidad" matemática aumenta de forma notable, la extensión del método del elemento finito a dos dimensiones es similar, conceptualmente, a los problemas unidimensionales analizados hasta ahora. De manera que se siguen los mismos pasos señalados en la sección 31.1.

31.3.1 Discretización

Comúnmente se emplean elementos sencillos, como triángulos o cuadriláteros, en la malla del elemento finito para dos dimensiones. En este análisis, nos limitaremos a elementos triangulares del tipo ilustrado en la figura 31.9.

31.3.2 Ecuaciones del elemento

Tal como en el caso unidimensional, el siguiente paso consiste en desarrollar una ecuación para aproximar la solución del elemento. Para un elemento triangular, la aproximación más sencilla es el polinomio lineal [compare con la ecuación (31.1)]

$$u(x, y) = a_0 + a_{1,1}x + a_{1,2}y \tag{31.28}$$

donde $u(x, y)$ = la variable dependiente, las a = coeficientes, x y y = variables independientes. Esta función debe pasar a través de los valores de $u(x, y)$ en los nodos del triángulo (x_1, y_1), (x_2, y_2) y (x_3, y_3). Por lo tanto,

$$u_1(x, y) = a_0 + a_{1,1}x_1 + a_{1,2}y_1$$

$$u_2(x, y) = a_0 + a_{1,1}x_2 + a_{1,2}y_2$$

$$u_3(x, y) = a_0 + a_{1,1}x_3 + a_{1,2}y_3$$

o, en forma matricial,

$$\begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_{1,1} \\ a_{1,2} \end{Bmatrix} = \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix}$$

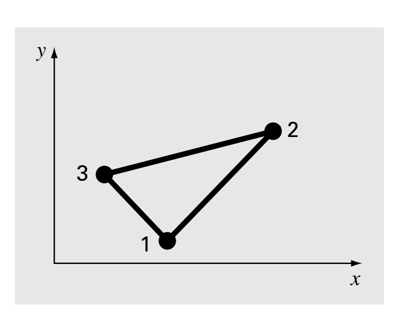


Figura 31.9
Un elemento triangular.

de donde se obtiene

$$a_0 = \frac{1}{2A_e}[u_1(x_2y_3 - x_3y_2) + u_2(x_3y_1 - x_1y_3) + u_3(x_1y_2 - x_2y_1)] \tag{31.29}$$

$$a_{1,1} = \frac{1}{2A_e}[u_1(y_2 - y_3) + u_2(y_3 - y_1) + u_3(y_1 - y_2)] \tag{31.30}$$

$$a_{1,2} = \frac{1}{2A_e}[u_1(x_3 - x_2) + u_2(x_1 - x_3) + u_3(x_2 - x_1)] \tag{31.31}$$

donde A_e es el área del elemento triangular,

$$A_e = \frac{1}{2}[(x_2y_3 - x_3y_2) + (x_3y_1 - x_1y_3) + (x_1y_2 - x_2y_1)]$$

Las ecuaciones (31.29) a (31.31) se sustituyen en la ecuación (31.28). Después de reagrupar términos semejantes, el resultado se expresa como sigue:

$$u = N_1u_1 + N_2u_2 + N_3u_3 \tag{31.32}$$

donde

$$N_1 = \frac{1}{2A_e}[(x_2y_3 - x_3y_2) + (y_2 - y_3)x + (x_3 - x_2)y]$$

$$N_2 = \frac{1}{2A_e}[(x_3y_1 - x_1y_3) + (y_3 - y_1)x + (x_1 - x_3)y]$$

$$N_3 = \frac{1}{2A_e}[(x_1y_2 - x_2y_1) + (y_1 - y_2)x + (x_2 - x_1)y]$$

La ecuación (31.32) permite predecir valores intermedios en el elemento, con base en los valores de sus nodos. En la figura 31.10 se muestra la función de forma junto con las funciones de interpolación correspondientes. Observe que la suma de las funciones de interpolación es siempre igual a 1.

Como en el caso unidimensional, hay varios métodos para desarrollar las ecuaciones del elemento, basados en la EDP y en las funciones de aproximación. Las ecuaciones resultantes son considerablemente más complicadas que la ecuación (31.26). Sin embargo, como las funciones de aproximación son normalmente polinomios de grado inferior como la ecuación (31.28), los términos de la matriz final del elemento consistirán de polinomios de grado inferior y de constantes.

31.3.3 Condiciones de la frontera y ensamble

La incorporación de condiciones de la frontera y el ensamble de la matriz del sistema también se hacen un poco más complicados cuando la técnica del elemento finito se aplique a problemas en dos y tres dimensiones. Sin embargo, como en la deducción de la matriz del elemento, la dificultad está más relacionada con la mecánica del proceso que con la complejidad conceptual. Por ejemplo, el establecimiento de la topología del sistema, que fue trivial para el caso unidimensional, se convierte en un asunto de gran importancia en los casos de dos y tres dimensiones. En particular, la elección de un esquema de numeración determinará el bandeado de la matriz del sistema resultante y, por lo tanto, la eficiencia con la que puede resolverse. En la figura 31.11 se muestra el esquema desarrollado antes para una placa caliente, y que se resolvió con los métodos por diferencias finitas en el capítulo 29.

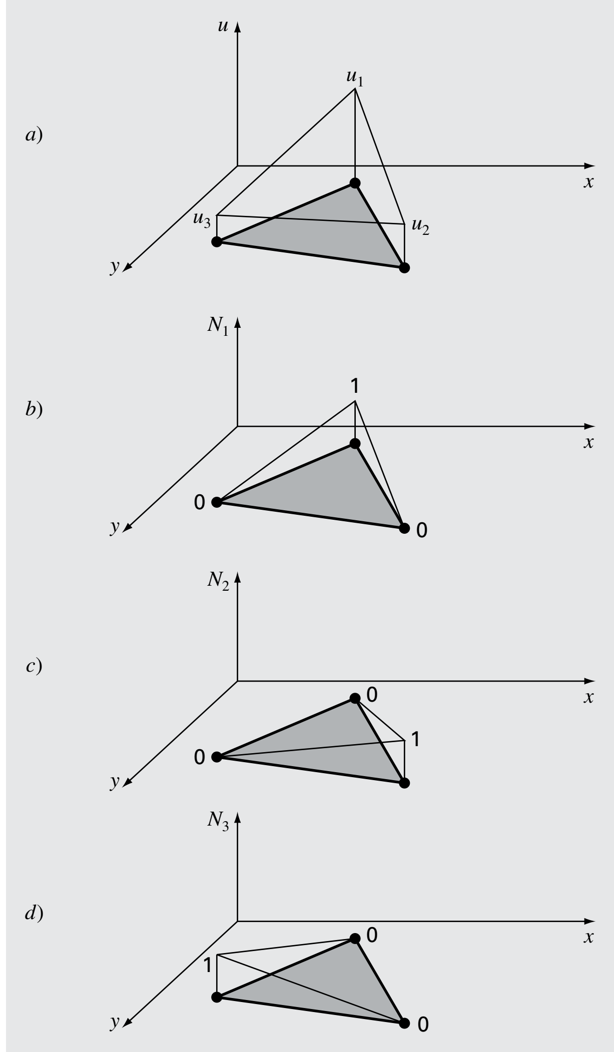


Figura 31.10
a) Una función de aproximación lineal para un elemento triangular. Las funciones de interpolación correspondientes se muestran en los incisos *b*) a *d*).

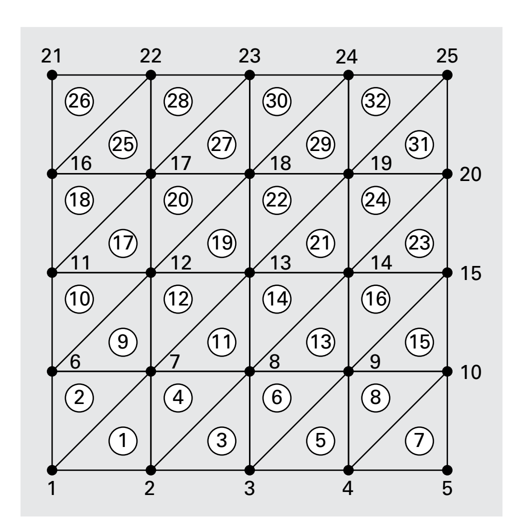


Figura 31.11
El esquema de numeración de los nodos y los elementos de una aproximación por elemento finito de la placa caliente, que se caracterizó previamente por diferencias finitas en el capítulo 29.

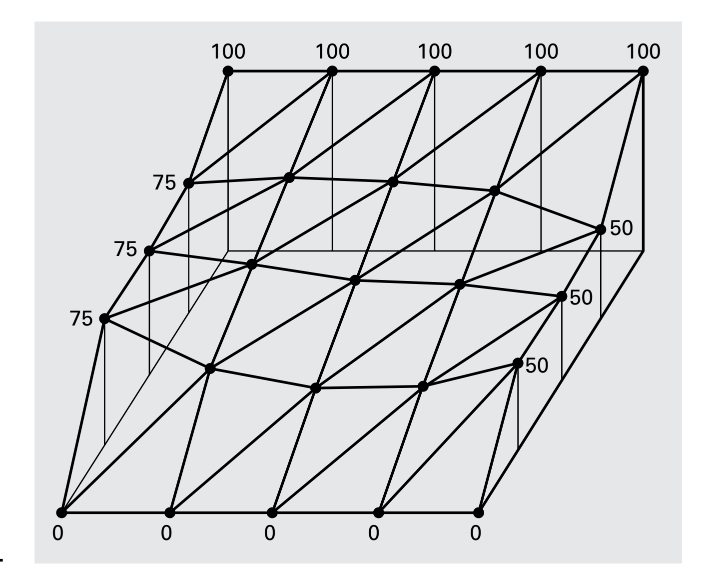


Figura 31.12
Distribución de temperatura en una placa caliente, calculada con el método del elemento finito.

31.3.4 Solución y procesamiento posterior

Aunque los mecanismos de solución son complicados, la matriz del sistema es tan sólo un conjunto de *n* ecuaciones simultáneas que pueden usarse para encontrar los valores de la variable dependiente en los *n* nodos. En la figura 31.12 se muestra una solución que corresponde a la solución por diferencias finitas de la figura 29.5.

31.4 RESOLUCIÓN DE EDP CON PAQUETES DE SOFTWARE

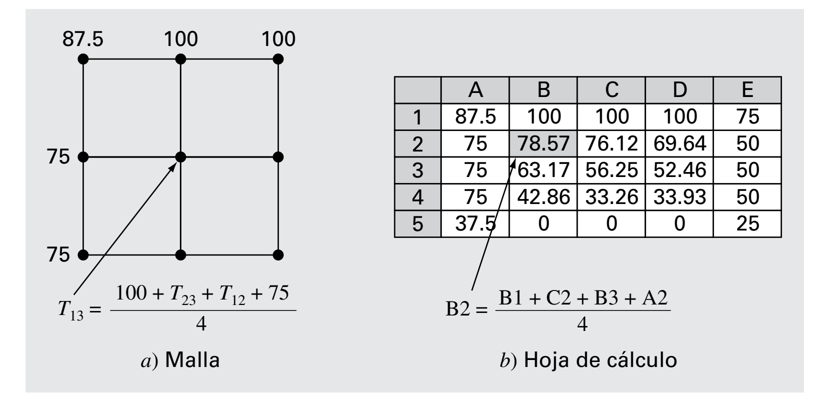
Los paquetes de software pueden ayudarnos a resolver directamente las EDP. Sin embargo, como se describe en las siguientes secciones, muchas de las soluciones están limitadas a problemas sencillos, lo cual es particularmente cierto para los casos de dos y de tres dimensiones. En tales situaciones, los paquetes genéricos (es decir, los no desarrollados expresamente para resolver EDP, como los paquetes para el elemento finito) a menudo están limitados a simples dominios rectangulares.

Aunque esto podría parecer una limitante, los problemas sencillos llegan a tener gran utilidad desde el punto de vista pedagógico. Ocurre así cuando las herramientas de visualización de los paquetes se utilizan para desplegar los resultados de los cálculos.

31.4.1 Excel

Aunque Excel no tiene la posibilidad de resolver directamente EDP, es un buen ambiente para desarrollar soluciones sencillas para las EDP elípticas. Por ejemplo, la presentación ortogonal de las celdas de la hoja de cálculo (figura 31.13*b*) es análoga a la malla utilizada en el capítulo 29 para modelar la placa caliente (figura 31.13*a*).

Como en la figura 31.13*b*, las condiciones de frontera de Dirichlet pueden introducirse primero en el contorno del bloque de la celda. La fórmula del método de Liebmann se implementa al introducir la ecuación (29.12) en una de las celdas del interior (como la celda B2 de la figura 31.13*b*). Así, el valor de la celda se calcula en función de las celdas adyacentes. Luego se copia la celda a las otras celdas interiores. Debido a la naturaleza relativa de la instrucción copiar de Excel, todas las demás celdas serán dependientes de sus celdas adyacentes.



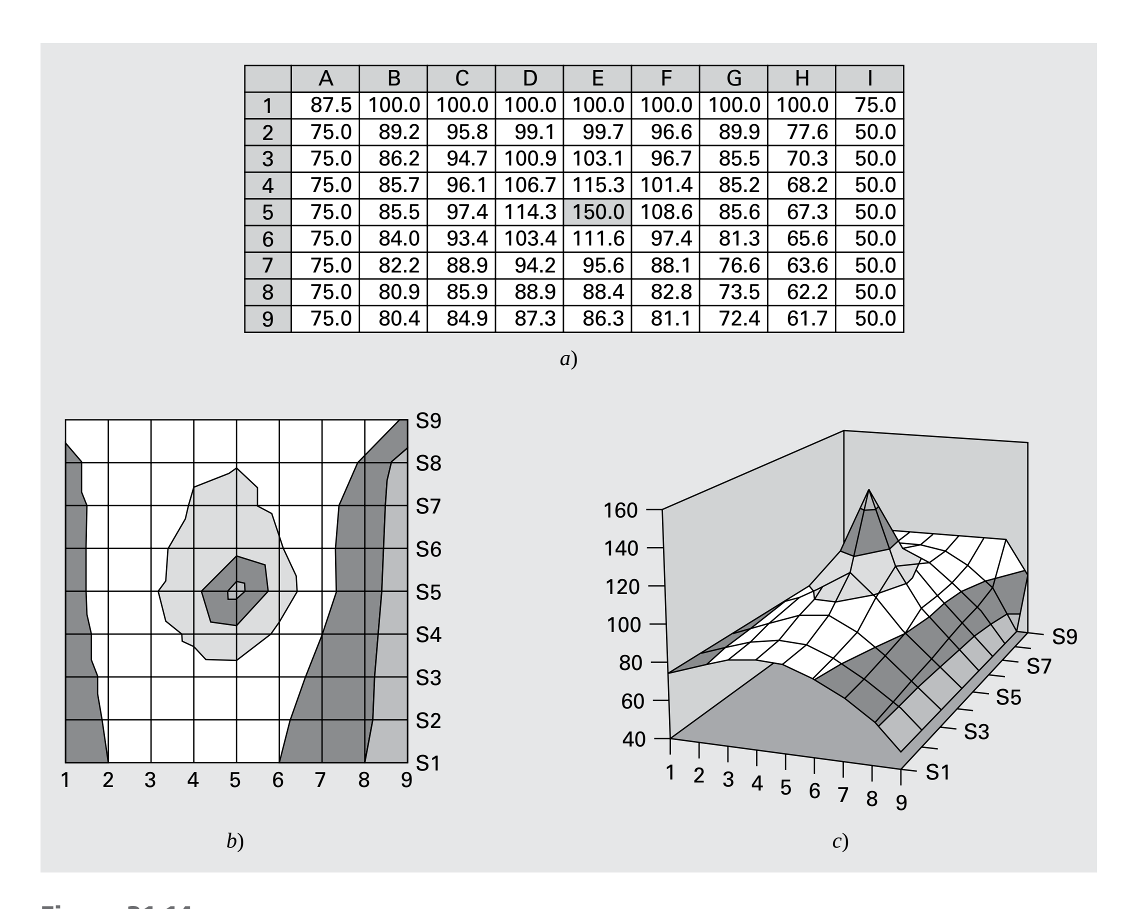
	A	B	C	D	E
1	87.5	100	100	100	75
2	75	78.57	76.12	69.64	50
3	75	63.17	56.25	52.46	50
4	75	42.86	33.26	33.93	50
5	37.5	0	0	0	25

Figura 31.13
Analogía entre *a*) una malla rectangular y *b*) las celdas de una hoja de cálculo.

Una vez que usted ha copiado la fórmula, probablemente obtendrá un mensaje de error: **Cannot resolve circular references** (No se pueden resolver referencias circulares). Usted puede rectificar esto al seleccionar **File** (**Archivo**), **Options** (**Opciones**) y haciendo clic en la categoría **Formulas** (**Fórmulas**). Luego seleccione **Calculation options** (**Opciones de cálculo**) y verifique el cuadro **Iterative calculation** (**Cálculo iterativo**). Esto permite que la hoja de cálculo vuelva a calcular (por omisión, son 100 iteraciones) y seguir el método de Liebmann iterativamente. Después de esto, presione la tecla F9 para volver a calcular de forma manual la hoja hasta que las respuestas no varíen, lo cual significa que ha convergido la solución.

Una vez resuelto el problema, se utilizan las herramientas gráficas de Excel para visualizar los resultados. En la figura 31.14*a* se muestra un ejemplo. En tal caso, se tiene que

- Se usó una malla fina
- Se aisló la frontera inferior
- Se agregó una fuente de calor de 150 a la mitad de la placa (celda E5).



	A	B	C	D	E	F	G	H	I
1	87.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	75.0
2	75.0	89.2	95.8	99.1	99.7	96.6	89.9	77.6	50.0
3	75.0	86.2	94.7	100.9	103.1	96.7	85.5	70.3	50.0
4	75.0	85.7	96.1	106.7	115.3	101.4	85.2	68.2	50.0
5	75.0	85.5	97.4	114.3	150.0	108.6	85.6	67.3	50.0
6	75.0	84.0	93.4	103.4	111.6	97.4	81.3	65.6	50.0
7	75.0	82.2	88.9	94.2	95.6	88.1	76.6	63.6	50.0
8	75.0	80.9	85.9	88.9	88.4	82.8	73.5	62.2	50.0
9	75.0	80.4	84.9	87.3	86.3	81.1	72.4	61.7	50.0

Figura 31.14
a) Solución en Excel de la ecuación de Poisson para una placa con un extremo inferior aislado y una fuente de calor. *b*) "Mapa topográfico" y *c*) ilustración tridimensional de las temperaturas.

Los resultados numéricos de la figura 31.14*a* pueden ilustrarse con el asistente para gráficos de Excel. Las figuras 31.14*b* y *c* muestran gráficas de superficies tridimensionales. Por lo general, la orientación *y* de éstas es la inversa de la hoja de cálculo. Así, el extremo superior de las temperaturas más altas (100) normalmente se representará en la parte inferior de la gráfica. Hemos invertido los valores de *y* en nuestra hoja antes de graficar, de modo que las gráficas sean consistentes con la hoja de cálculo.

Advierta cómo las gráficas nos ayudan a visualizar lo que sucederá. El calor fluye hacia abajo desde la fuente hasta las fronteras, formando una imagen parecida a una montaña. El calor también fluye hacia abajo desde la frontera con temperatura alta hasta los dos extremos laterales. Observe cómo el calor fluye hacia el extremo de baja temperatura (50). Por último, observe cómo el gradiente de temperatura en la dimensión *y* tiende a cero para el extremo inferior aislado ($\partial T/\partial y \to 0$).

31.4.2 MATLAB

Aunque el paquete MATLAB estándar no tiene grandes capacidades para resolver las EDP, se pueden desarrollar archivos M y funciones con este propósito. Además, su capacidad para mostrar imágenes es muy útil, en particular para visualizar problemas bidimensionales.

Para ilustrar esta capacidad, primero desarrollamos la hoja de cálculo en Excel de la figura 31.14*a*. Estos resultados pueden guardarse como un archivo de texto, con un nombre como **plate.txt**. Después este archivo se traslada al directorio de MATLAB.

Una vez en MATLAB, se carga este archivo escribiendo

```
>> load plate.txt
```

Luego, los gradientes se calculan simplemente así

```
>> [px,py]=gradient(plate);
```

Observe que éste es el método más simple para calcular gradientes usando los valores por omisión de dx = dy = 1. Por lo tanto, serán correctas las direcciones y las magnitudes relativas.

Por último, se utilizan una serie de comandos para obtener la gráfica. El comando `contour` desarrolla una gráfica de contorno de los datos. El comando `clabel` agrega etiquetas de contorno a la gráfica. Finalmente, `quiver` toma los datos del gradiente y los añade a la gráfica en forma de flechas.

```
>> cs=contour(plate);clabel(cs);hold on
>> quiver(-px,-py);hold off
```

Observe que se ha agregado el signo menos, debido al signo menos de la ley de Fourier [ecuación (29.4)]. Como se ve en la figura 31.15, la gráfica resultante proporciona una excelente representación de la solución.

Observe que cualquier archivo en el formato adecuado se puede introducir en MATLAB y se puede mostrar de esta manera. Esta posibilidad de compartir archivos entre las herramientas se está volviendo usual. Además, se pueden crear archivos en una ubicación en una herramienta, transmitirlos por internet a otra ubicación donde el archivo se puede abrir con otra herramienta. Éste es uno de los aspectos atractivos de las aplicaciones numéricas modernas.

31.4.3 Mathcad

Mathcad tiene dos funciones que pueden resolver la ecuación de Poisson. Usted puede usar la función **relax** cuando sepa el valor de la incógnita en los cuatro lados de una región cuadrada. Esta función resuelve un sistema de ecuaciones algebraicas lineales usando la iteración de Gauss-Seidel con sobrerrelajamiento para acelerar la rapidez de convergencia. Para el caso especial en el que hay fuentes internas o receptores internos y la incógnita es cero en todos los cuatro lados del cuadrado, entonces usted puede usar la función **multigrid**, que es usualmente más rápida que la función **relax**. Ambas funciones devuelven una matriz cuadrada donde la ubicación del elemento corresponde a su ubicación dentro de la región cuadrada. El valor del elemento se aproxima al valor de la solución de la ecuación de Poisson en este punto.

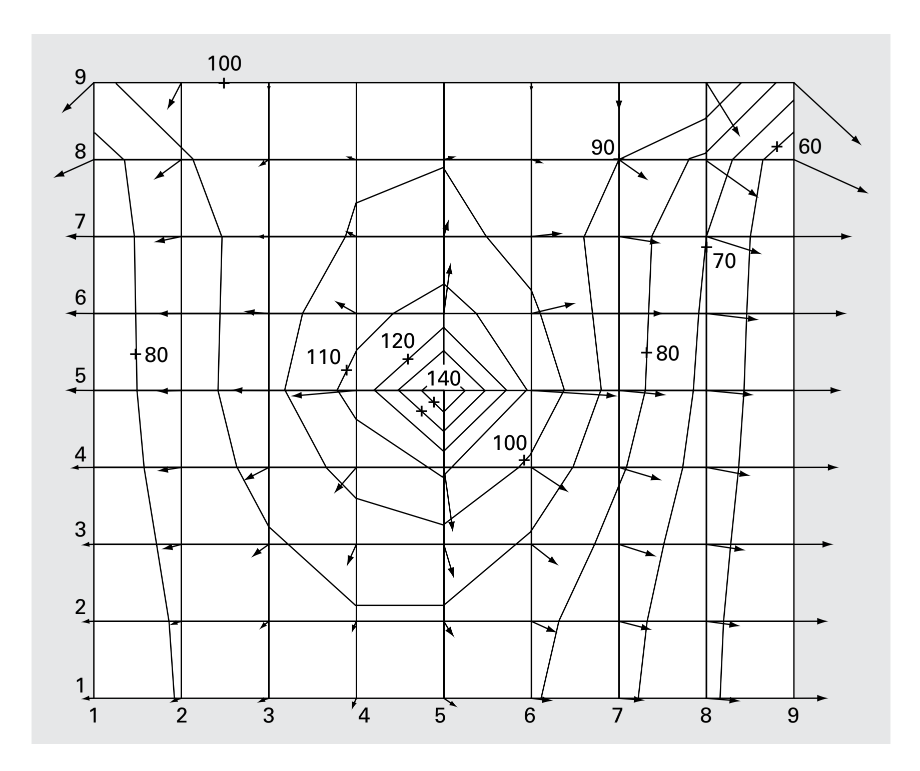


Figura 31.15
Gráficas de contorno generadas en MATLAB y calculadas con Excel, para la placa caliente (figura 31.14).

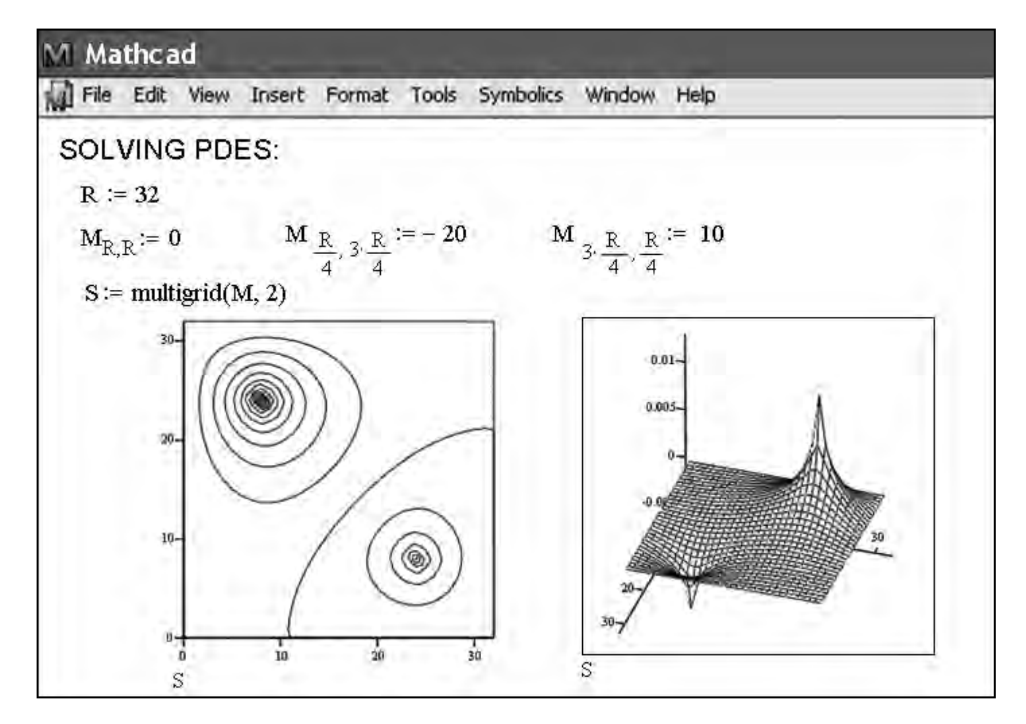


Figura 31.16
Pantalla de Mathcad para determinar la solución de una EDP elíptica.

SOFTWARE

La figura 31.16 muestra un ejemplo en el que una placa cuadrada contiene fuentes de calor mientras la frontera se mantiene en cero. El primer paso es establecer dimensiones para la malla de temperatura y la matriz de la fuente de calor. La cuadrícula de temperatura tiene dimensiones (R + 1) × (R + 1), mientras que la matriz de la fuente de calor es R × R. Por ejemplo, una cuadrícula de temperatura de 3 × 3 tiene 4 (2 × 2) fuentes de calor posibles. En este caso, se establece una cuadrícula de temperatura de 33 × 33 y una matriz de fuente de calor de 32 × 32. El comando Mathcad M_{RR}:= 0 (con R = 32) establece las dimensiones de la matriz de fuentes y pone en cero todos los elementos. En seguida se establecen la ubicación y la fuerza de dos fuentes de calor. Finalmente, *S* es la distribución resultante de temperatura calculada con la función **multigrid**. El segundo argumento de **multigrid** es un parámetro que controla la exactitud numérica. Como lo sugiere la asistencia de Mathcad, un valor de 2 generalmente da una buena aproximación de la solución.

La distribución de temperatura se puede mostrar con gráficas de superficie, contorno o campo vectorial. Éstas se pueden colocar en cualquier sitio en la hoja de trabajo haciendo clic en la ubicación deseada. Esto dispone una retícula roja en la ubicación. Luego use el menú desplegables Insert/Graph para colocar una gráfica vacía en la hoja de cálculo con espacios para las expresiones que se van a graficar y para los rangos de las variables. Simplemente teclee *S* en el espacio sobre el eje *z*. Mathcad hace el resto para producir las gráficas que se muestran en la figura 31.16. Una vez que se ha creado la gráfica, usted puede usar los menús desplegables Format/Surface Plot y Format/Contour Plot para cambiar el color o añadir títulos, rótulos y otras características.

PROBLEMAS

31.1 Repita el ejemplo 31.1, pero para $T(0, t) = 75$ y $T(10, t) = 150$, y una fuente uniforme de calor de 15.

31.2 Repita el ejemplo 31.2, pero para condiciones de frontera de $T(0, t) = 75$ y $T(10, t) = 150$, y una fuente de calor de 15.

31.3 Aplique los resultados del problema 31.2 para calcular la distribución de temperatura para la barra completa, con el uso del enfoque del elemento finito.

31.4 Utilice el método de Galerkin para desarrollar una ecuación de elemento para una versión de estado estable de la ecuación de advección-difusión descrita en el problema 30.7. Exprese el resultado final en el formato de la ecuación (31.26) de modo que cada término tenga una interpretación física.

31.5 El modelo siguiente es una versión de la ecuación de Poisson que ocurre en la mecánica para la deflexión vertical de una barra con una carga distribuida $P(x)$:

$$A_c E \frac{\partial^2 u}{\partial x^2} = P(x)$$

donde A_c = área de la sección transversal, E = módulo de Young, u = deflexión y x = distancia medida a lo largo de la longitud de la barra. Si la barra está fija rígidamente ($u = 0$) por ambos extremos, use el método del elemento finito para modelar sus deflexiones para A_c = 0.1 m², $E = 200 \times 10^9$ N/m², L = 10 m y $P(x)$ = 1 000 N/m. Emplee un valor de Δx = 2 m.

31.6 Desarrolle un programa amigable para el usuario a fin de modelar la distribución de estado estable de la temperatura en una barra con fuente de calor constante, con el método del elemento finito. Elabore el programa de modo que se utilicen nodos irregularmente espaciados.

31.7 Utilice Excel para realizar el mismo cálculo que en la figura 31.14, pero aísle el borde del lado derecho y agregue una fuente de calor de –150 en la celda C7.

31.8 Emplee MATLAB o Mathcad para desarrollar una gráfica de contorno con flechas de flujo para la solución en Excel del problema 31.7.

31.9 Use Excel para modelar la distribución de temperatura de la placa que se muestra en la figura P31.9. La placa tiene 0.02 m de espesor y una conductividad térmica de 3 W/(m · °C).

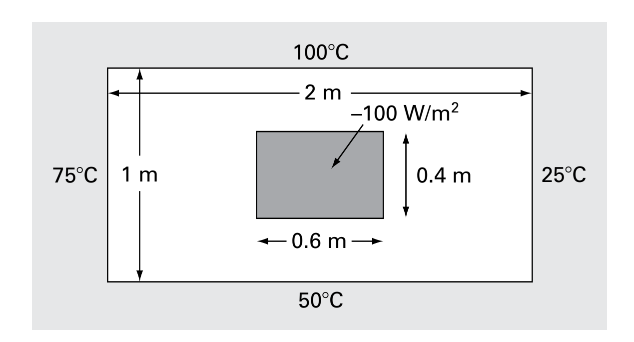


Figura P31.9

31.10 Use MATLAB o Mathcad para desarrollar una gráfica de contorno con flechas de flujo para la solución con Excel del problema 31.9.

31.11 Encuentre la distribución de temperatura en una barra (véase la figura P31.11) con generación de calor interno, por medio del método del elemento finito. Obtenga las ecuaciones nodales de elemento con el uso de la conducción de calor de Fourier.

$$q_k = -kA \frac{dT}{\partial x}$$

y las relaciones de conservación del calor

$$\sum \; [q_k + f(x)] = 0$$

donde q_k = flujo de calor (W), k = conductividad térmica (W/(m · °C)), A = área de la sección transversal (m²) y $f(x)$ = fuente de calor (W/

cm). La barra tiene un valor de kA = 100 W m/°C. La barra mide 50 cm de largo, en el extremo izquierdo la coordenada x es cero y positiva hacia la derecha. Divida la barra en 5 elementos (6 nodos, cada elemento de 10 cm de largo). El extremo izquierdo de la barra tiene un gradiente fijo de temperatura y la temperatura es variable. El extremo derecho tiene una temperatura fija y el gradiente es variable. La fuente de calor $f(x)$ tiene un valor constante. Así, las condiciones son

$$\left.\frac{dT}{\partial x}\right|_{x=0} = 0.25°\text{C/m} \quad T|_{x=50} = 100°\text{C} \quad f(x) = 30 \text{ W/cm}$$

Desarrolle las ecuaciones nodales que deben resolverse para las temperaturas y los gradientes de temperatura en cada uno de los seis nodos. Acople las ecuaciones, inserte las condiciones de frontera y resuelva el conjunto resultante para las incógnitas.

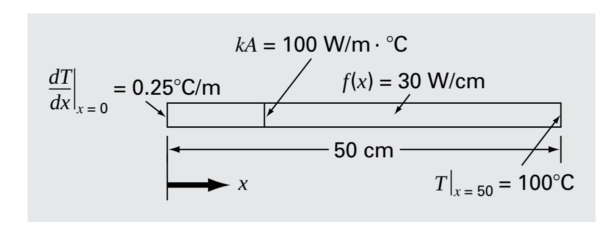


Figura P31.11

31.12 Encuentre la distribución de temperatura en una barra (véase la figura P31.12) con generación interna de calor, con el método del elemento finito. Obtenga las ecuaciones de elementos nodales con el uso de la conducción de calor de Fourier.

$$q_k = -kA\frac{dT}{\partial x}$$

y las relaciones de conservación de calor

$$\sum [q_k + f(x)] = 0$$

donde q_k = flujo de calor (W), k = conductividad térmica (W/(m · °C)), A = área de la sección transversal (m^2) y $f(x)$ = fuente de calor (W/cm). La barra mide 50 cm de longitud, la coordenada x es cero en el extremo izquierdo y positiva hacia la derecha. La barra también tiene bloqueos lineales en x = 0 y en x = 50, con un valor de kA = 100 y 50 W m/°C, respectivamente. Divida la barra en 5 elementos (6 nodos, cada elemento de 10 cm de largo). Ambos extremos de la barra tienen temperaturas fijas. La fuente de calor $f(x)$ tiene un valor constante. Así, las condiciones son

$$T|_{x=0} = 100°\text{C} \quad T|_{x=50} = 50°\text{C} \quad f(x) = 30 \text{ W/cm}$$

Las áreas bloqueadas deben tratarse como si fueran constantes en la longitud de un elemento. Por lo tanto, promedie los valores kA en cada extremo del nodo y tome el promedio como una constante en el nodo. Desarrolle las ecuaciones de elementos nodales que deben resolverse para las temperaturas y los gradientes de temperatura en cada uno de los seis nodos. Ensamble las ecuaciones, inserte las condiciones de frontera y resuelva para el conjunto resultante de incógnitas.

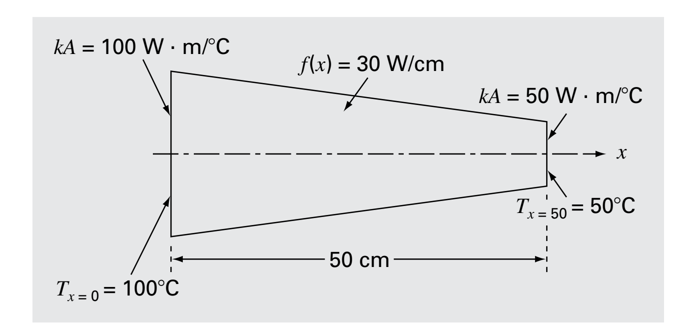


Figura P31.12

31.13 Use un paquete de software para resolver la distribución de temperatura de la placa en forma de L de la figura P29.18. Muestre sus resultados como gráfica de contorno con flechas de flujo.

CAPÍTULO 32

Estudio de casos: ecuaciones diferenciales parciales

El propósito de este capítulo es aplicar los métodos de la parte ocho a problemas prácticos de ingeniería. En la *sección 32.1* se utiliza una EDP parabólica para calcular la distribución de una sustancia química, dependiente del tiempo a lo largo del eje longitudinal de un reactor rectangular. Este ejemplo ilustra cómo la inestabilidad de una solución puede deberse a la naturaleza de la EDP, más que a las propiedades del método numérico.

Las secciones 32.2 y 32.3 presentan aplicaciones de las ecuaciones de Poisson y Laplace a problemas de ingeniería civil y eléctrica. Entre otras cuestiones, esto le permitirá distinguir tanto las similitudes como las diferencias entre los problemas en esas áreas de la ingeniería. Además, se pueden comparar con el problema de la placa caliente que ha servido como sistema prototipo en esta parte del libro. La *sección 32.2* trata de la deflexión de una placa cuadrada; mientras que la *sección 32.3* se dedica al cálculo de la distribución del voltaje y el flujo de carga en una superficie bidimensional con un extremo curvado.

La *sección 32.4* presenta un análisis del elemento finito aplicado a una serie de resortes. Este problema de mecánica y estructuras ilustra mejor las aplicaciones del elemento finito, que al problema de temperatura usado para analizar el método en el capítulo 31.

32.1 BALANCE DE MASA UNIDIMENSIONAL DE UN REACTOR (INGENIERÍA QUÍMICA/BIOINGENIERÍA)

Antecedentes Los ingenieros químicos utilizan mucho los reactores idealizados en su trabajo de diseño. En las secciones 12.1 y 28.1 nos concentramos en reactores simples o acoplados bien mezclados, los cuales constituyen ejemplos de *sistemas de parámetros localizados* (recuerde la sección PT3.1.2).

La figura 32.1 muestra un reactor alargado con una sola entrada y una salida. Este reactor puede caracterizarse como un *sistema de parámetros distribuidos*. Si se supone que la sustancia química que va a modelarse está sujeta a un decaimiento[1] de primer orden, y el tanque está bien mezclado vertical y lateralmente, se realiza un balance de masa en un segmento finito de longitud Δx, como sigue

[1] Es decir, la sustancia química decae a una velocidad que es linealmente proporcional a la cantidad de sustancia química presente.

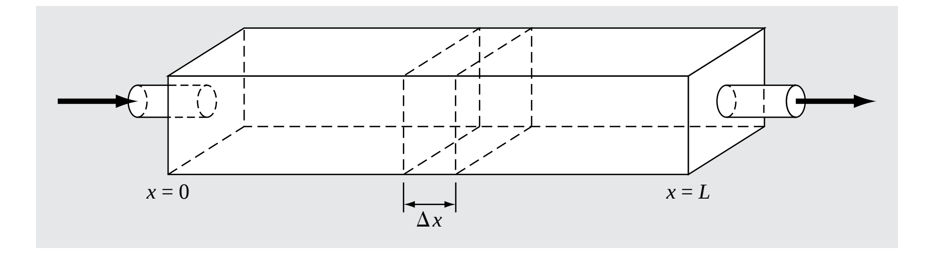


Figura 32.1
Reactor alargado con un solo punto de entrada y salida. Un balance de masa se desarrolla alrededor de un segmento finito a lo largo del eje longitudinal del tanque con el objetivo de deducir una ecuación diferencial para la concentración.

$$V\frac{\Delta c}{\Delta t} = \underbrace{Qc(x)}_{\text{Flujo de entrada}} \underbrace{-Q\left[c(x)+\frac{\partial c(x)}{\partial x}\Delta x\right]}_{\text{Flujo de salida}} - \underbrace{DA_c\frac{\partial c(x)}{\partial x}}_{\text{Dispersión a la entrada}}$$

$$\underbrace{+DA_c\left[\frac{\partial c(x)}{\partial x}+\frac{\partial}{\partial x}\frac{\partial c(x)}{\partial x}\Delta x\right]}_{\text{Dispersión a la salida}} - \underbrace{kVc}_{\text{Reacción de decaimiento}} \qquad (32.1)$$

donde V = volumen (m^3), Q = flujo volumétrico (m^3/h), c = concentración ($moles/m^3$), D es un coeficiente de dispersión (m^2/h), A_c es el área de la sección transversal del reactor (m^2) y k es el coeficiente de decaimiento de primer orden (h^{-1}). Observe que los términos de dispersión están basados en la *primera ley de Fick*,

$$\text{Flujo} = -D\frac{\partial c}{\partial x} \qquad (32.2)$$

que es análoga a la ley de Fourier para la conducción del calor [recuerde la ecuación (29.4)]. Esta ecuación especifica que la turbulencia de mezclado tiende a mover la masa desde regiones de alta hasta las de baja concentración. El parámetro D, por lo tanto, determina la magnitud de la turbulencia de mezclado.

Si Δx y Δt tienden a cero, la ecuación (32.1) será

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} - U\frac{\partial c}{\partial x} - kc \qquad (32.3)$$

donde $U = Q/A_c$ es la velocidad del agua que fluye a través del reactor. El balance de masa de la figura 32.1, por lo tanto, se expresa ahora como una ecuación diferencial parcial parabólica. En ocasiones, a la ecuación (32.3) se le llama la *ecuación de advección-dispersión* con reacción de primer orden. En estado estacionario, se reduce a una EDO de segundo orden,

$$0 = D\frac{d^2 c}{dx^2} - U\frac{d^2 c}{dx} - kc \qquad (32.4)$$

Antes de $t = 0$, el reactor se llena con agua libre de la sustancia química. En $t = 0$, se inyecta la sustancia química en el flujo de entrada del reactor a un nivel constante de c_{en}. Así, se tienen las siguientes condiciones de frontera:

$$Qc_{en} = Qc_0 - DA_c\frac{\partial c_0}{\partial x}$$

y

$$c'(L, t) = 0$$

La segunda condición especifica que la sustancia sale del reactor simplemente como una función del flujo a través del tubo de salida. Es decir, se supone que la dispersión en el reactor no afecta la velocidad de salida. Bajo estas condiciones, utilice los métodos numéricos para resolver la ecuación (32.4) en niveles de estado estacionario para el reactor. Observe que se trata de un problema de EDO con valores en la frontera. Después resuelva la ecuación (32.3) para caracterizar la respuesta transitiva (es decir, cómo cambian los niveles con el tiempo conforme el sistema se aproxima al estado estacionario). Esta aplicación utiliza una EDP.

Solución Se desarrolla una ecuación en estado estacionario sustituyendo la primera y la segunda derivadas de la ecuación (32.4) por diferencias finitas centradas para obtener

$$0 = D\frac{c_{i+1} - 2c_i + c_{i-1}}{x^2} - U\frac{c_{i+1} - c_{i-1}}{2\,x} - kc_i$$

Agrupando términos se tiene

$$-\left(\frac{D}{U\,\Delta x} + \frac{1}{2}\right)c_{i-1} + \left(\frac{2D}{U\,\Delta x} + \frac{k\,\Delta x}{U}\right)c_0 - \left(\frac{D}{U\,\Delta x} - \frac{1}{2}\right)c_{i+1} = 0 \tag{32.5}$$

Esta ecuación se puede dar para cada uno de los nodos del sistema. En los extremos del reactor, este proceso introduce nodos que están fuera del sistema. Por ejemplo, en el nodo de entrada ($i = 0$),

$$-\left(\frac{D}{U\,\Delta x} + \frac{1}{2}\right)c_{-1} + \left(\frac{2D}{U\,\Delta x} + \frac{k\,\Delta x}{U}\right)c_0 - \left(\frac{D}{U\,\Delta x} - \frac{1}{2}\right)c_1 = 0 \tag{32.6}$$

El término c_{-1} se elimina utilizando la primera condición de frontera. A la entrada, se debe satisfacer el siguiente balance de masa:

$$Qc_{\text{en}} = Qc_0 - DA_c\frac{\partial c_0}{\partial x}$$

donde c_0 = concentración en $x = 0$. Así, esta condición de frontera especifica que la cantidad de sustancia química transportada hacia el tanque por advección a través del tubo debe ser igual a la cantidad llevada hacia afuera desde la entrada, tanto por advección como por dispersión de turbulencia en el reactor. Se sustituye la derivada por una diferencia dividida finita

$$Qc_{\text{en}} = Qc_0 - DA_c\frac{c_1 - c_{-1}}{2\,\Delta x}$$

De la cual se despeja c_{-1}

$$c_{-1} = c_1 + \frac{2\,\Delta x\,U}{D}c_{\text{en}} - \frac{2\,\Delta x\,U}{D}c_0$$

que al sustituirse en la ecuación (32.6) se obtiene

$$\left(\frac{2D}{U\,\Delta x} + \frac{k\,\Delta x}{U} + 2 + \frac{\Delta xU}{D}\right)c_0 - \left(\frac{2D}{U\,\Delta x}\right)c_1 = \left(2 + \frac{\Delta xU}{D}\right)c_{\text{en}} \tag{32.7}$$

Se puede realizar un desarrollo similar para la salida, donde la ecuación en diferencias original es

$$-\left(\frac{D}{U\,\Delta x} + \frac{1}{2}\right)c_{n-1} + \left(\frac{2D}{U\,\Delta x} + \frac{k\,\Delta x}{U}\right)c_n - \left(\frac{D}{U\,\Delta x} - \frac{1}{2}\right)c_{n+1} = 0 \tag{32.8}$$

La condición de frontera a la salida es

$$Qc_n - DA_c\frac{dc_n}{dx} = Qc_n$$

Como en la entrada, se utiliza una diferencia dividida para aproximar la derivada,

$$Qc_n - DA_c\frac{c_{n+1} - c_{n-1}}{2\,\Delta x} = Qc_n \tag{32.9}$$

Una inspección de esta ecuación nos lleva a concluir que $c_{n+1} = c_{n-1}$. En otras palabras, la pendiente a la salida debe ser cero para que se satisfaga la ecuación (32.9). Sustituyendo este resultado en la ecuación (32.8) y simplificando, se tiene

$$-\left(\frac{2D}{U\,\Delta x}\right)c_{n-1} + \left(\frac{2D}{U\,\Delta x} + \frac{k\,\Delta x}{U}\right)c_n = 0 \qquad (32.10)$$

Las ecuaciones (32.5), (32.7) y (32.10) forman ahora un sistema tridiagonal de n ecuaciones con n incógnitas. Por ejemplo, si $D = 2$, $U = 1$, $\Delta x = 2.5$, $k = 0.2$ y $c_{en} = 100$, el sistema es

$$\begin{bmatrix} 5.35 & -1.6 & & & \\ -1.3 & 2.1 & -0.3 & & \\ & -1.3 & 2.1 & -0.3 & \\ & & -1.3 & 2.1 & -0.3 \\ & & & -1.6 & 2.1 \end{bmatrix} \begin{Bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \\ c_4 \end{Bmatrix} = \begin{Bmatrix} 325 \\ 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix}$$

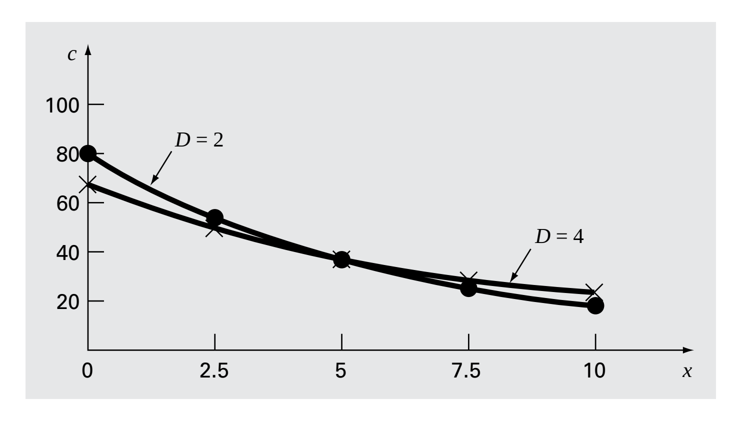


Figura 32.2
Concentración contra distancia a lo largo del eje longitudinal de un reactor rectangular para una sustancia química que decae con cinética de primer orden.

de donde se obtiene

$c_0 = 76.44 \qquad c_1 = 52.47 \qquad c_2 = 36.06 \qquad c_3 = 25.05 \qquad c_4 = 19.09$

La gráfica de estos resultados se muestra en la figura 32.2. Como se esperaba, la concentración disminuye debido a la reacción de decaimiento, conforme la sustancia química fluye a través del reactor. Además del cálculo anterior, la figura 32.2 muestra otro caso con $D = 4$. Observe cómo aumentando la turbulencia de mezclado la curva tiende a ser plana.

En cambio, si se disminuye la dispersión, la curva será más pronunciada conforme el mezclado sea menos importante en relación con la advección y el decaimiento. Debe notarse que si la dispersión disminuye demasiado, el cálculo estará sujeto a errores numéricos. Este tipo de error se conoce como *inestabilidad estática*, en contraste con la *inestabilidad dinámica* debida a largos lapsos de tiempo durante un cálculo dinámico. El criterio para evitar esta inestabilidad estática es

$$\Delta x \leq \frac{2D}{U}$$

Así, el criterio se vuelve más riguroso (con una Δx más baja) para los casos donde la advección domina sobre la dispersión.

Además de los cálculos en estado estacionario, los métodos numéricos se utilizan para generar la solución variable con el tiempo de la ecuación (32.3). La figura 32.3 muestra los resultados para $D = 2$, $U = 1$, $\Delta x = 2.5$, $k = 0.2$ y $c_{en} = 100$, donde la concentración en el tanque es 0 en el instante cero. Como se esperaba, el impacto inmediato ocurre cerca de la entrada. Con el tiempo, la solución se aproxima al nivel de estado estacionario.

Debe observarse que, en tales cálculos dinámicos, el lapso de tiempo está restringido por un criterio de estabilidad expresado como (Chapra, 1997)

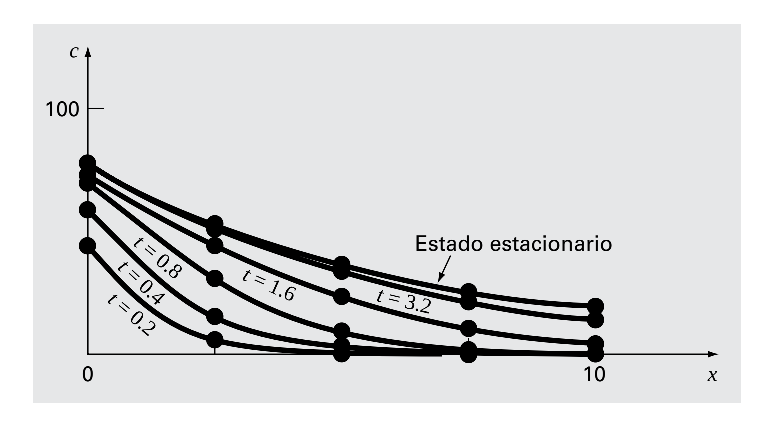


Figura 32.3
Concentración contra distancia a diferentes instantes, durante la acumulación de la sustancia química en un reactor.

$$\Delta t \leq \frac{(\Delta x)^2}{2D + k(\Delta x)^2}$$

Así, el término de la reacción actúa para hacer más pequeño el lapso de tiempo.

32.2 DEFLEXIONES DE UNA PLACA (INGENIERÍA CIVIL/AMBIENTAL)

Antecedentes Una placa cuadrada, apoyada simplemente en sus extremos está sujeta a una carga por unidad de área q (figura 32.4). La deflexión en la dimensión z se determina resolviendo la EDP elíptica (véase Carnahan, Luther y Wilkes, 1969)

$$\frac{\partial^4 z}{\partial x^4} + 2\frac{\partial^4 z}{\partial x^2 \partial y^2} + \frac{\partial^4 z}{\partial y^4} = \frac{q}{D} \tag{32.11}$$

sujeta a condiciones de frontera en los extremos, donde la deflexión y la pendiente normal a la frontera son cero. El parámetro D es la rigidez de flexión,

$$D = \frac{E\,\Delta z^3}{12(1-\sigma^2)} \tag{32.12}$$

donde E = el módulo de elasticidad, Δz = el espesor de la placa y σ = razón de Poisson.

Si definimos una nueva variable como sigue

$$u = \frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2}$$

La ecuación (32.11) se reexpresa como

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{q}{D} \tag{32.13}$$

Por lo tanto, el problema se reduce a resolver de manera sucesiva dos ecuaciones de Poisson. Primero, de la ecuación (32.13) se obtiene u sujeta a la condición de frontera $u = 0$ en los extremos. Después, los resultados se emplean junto con

$$\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = u \tag{32.14}$$

para obtener z sujeta a la condición de que $z = 0$ en los extremos.

Desarrolle un programa computacional para determinar las deflexiones de una placa cuadrada sujeta a una carga constante por unidad de área. Pruebe el programa con una placa de 2 metros de longitud en sus extremos, $q = 33.6$ kN/m^2, $\sigma = 0.3$, $\Delta z = 10^{-2}$ m y $E = 2 \times 10^{11}$ Pa. Emplee $\Delta x = \Delta y = 0.5$ m para su corrida de prueba.

Solución Las diferencias divididas finitas se emplean para sustituir cada uno de los sumandos en la ecuación (32.13) para dar

$$\frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{\Delta x^2} + \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{\Delta y^2} = \frac{q}{D} \tag{32.15}$$

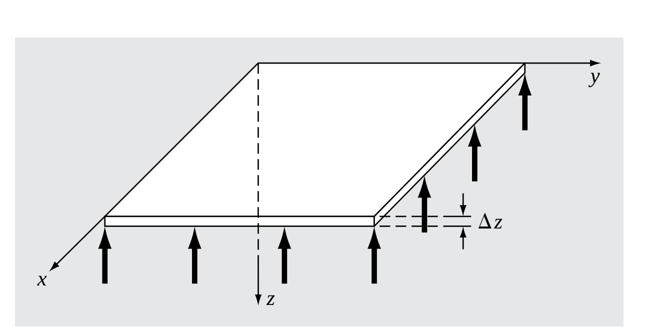


Figura 32.4
Placa cuadrada apoyada en forma simple, sujeta a una carga en la superficie.

La ecuación (32.12) se utiliza para calcular $D = 1.832 \times 10^4$ N/m. Este resultado, junto con los otros parámetros del sistema, se sustituyen en la ecuación (32.15) para obtener

$$u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{i,j} = 0.458$$

Esta ecuación se puede dar para todos los nodos con las fronteras en $u = 0$. Las ecuaciones resultantes son

$$\begin{bmatrix} -4 & 1 & & 1 & & & & & \\ 1 & -4 & 1 & & 1 & & & & \\ & 1 & -4 & & & 1 & & & \\ 1 & & & -4 & 1 & & 1 & & \\ & 1 & & 1 & -4 & 1 & & 1 & \\ & & 1 & & 1 & -4 & & & 1 \\ & & & 1 & & & -4 & 1 & \\ & & & & 1 & & 1 & -4 & 1 \\ & & & & & 1 & & 1 & -4 \end{bmatrix} \begin{Bmatrix} u_{1,1} \\ u_{2,1} \\ u_{3,1} \\ u_{1,2} \\ u_{2,2} \\ u_{3,2} \\ u_{1,3} \\ u_{2,3} \\ u_{3,3} \end{Bmatrix} = \begin{Bmatrix} 0.458 \\ 0.458 \\ 0.458 \\ 0.458 \\ 0.458 \\ 0.458 \\ 0.458 \\ 0.458 \\ 0.458 \end{Bmatrix}$$

de las cuales se obtiene

$u_{1,1} = -0.315$ $\quad u_{1,2} = -0.401$ $\quad u_{1,3} = -0.315$
$u_{2,1} = -0.401$ $\quad u_{2,2} = -0.515$ $\quad u_{2,3} = -0.401$
$u_{3,1} = -0.315$ $\quad u_{3,2} = -0.401$ $\quad u_{3,3} = -0.315$

Estos resultados, a su vez, se sustituyen en la ecuación (32.14), que se escribe en forma de diferencias finitas para obtener

$z_{1,1} = 0.063$ $\quad z_{1,2} = 0.086$ $\quad z_{1,3} = 0.063$
$z_{2,1} = 0.086$ $\quad z_{2,2} = 0.118$ $\quad z_{2,3} = 0.086$
$z_{3,1} = 0.063$ $\quad z_{3,2} = 0.086$ $\quad z_{3,3} = 0.063$

32.3 PROBLEMAS DE CAMPO ELECTROSTÁTICO BIDIMENSIONAL (INGENIERÍA ELÉCTRICA)

Antecedentes Así como la ley de Fourier y el balance de calor se emplean para caracterizar la distribución de temperatura, existen relaciones análogas para modelar problemas en otras áreas de la ingeniería. Por ejemplo, los ingenieros eléctricos usan un método similar cuando modelan campos electrostáticos.

Bajo varias suposiciones de simplificación, un análogo de la ley de Fourier se representa en forma unidimensional como

$$D = -\varepsilon \frac{dV}{dx}$$

donde D se conoce como el vector de densidad de flujo eléctrico, ε = permitividad eléctrica del material y V = potencial electrostático.

De manera similar, una ecuación de Poisson para campos electrostáticos se representa en dos dimensiones de la siguiente manera

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = -\frac{\rho_v}{\varepsilon} \tag{32.16}$$

donde ρ_v = densidad de carga volumétrica.

Por último, en regiones que no contienen carga libre (es decir, $\rho_v = 0$), la ecuación (32.16) se reduce a la ecuación de Laplace,

$$\frac{\partial^2 V}{\partial x^2}+\frac{\partial^2 V}{\partial y^2}=0 \tag{32.17}$$

Emplee métodos numéricos para resolver la ecuación (32.17) para la situación mostrada en la figura 32.5. Calcule los valores para V y para D si $\varepsilon = 2$.

Solución Usando el procedimiento que se describe en la sección 29.3.2, la ecuación (29.24) para el nodo (1, 1) se escribe como

$$\frac{2}{\Delta x^2}\left[\frac{V_{1,1}-V_{0,1}}{\alpha_1(\alpha_1+\alpha_2)}+\frac{V_{1,1}-V_{2,1}}{\alpha_2(\alpha_1+\alpha_2)}\right]+\frac{2}{\Delta y^2}\left[\frac{V_{1,1}-V_{0,1}}{\beta_1(\beta_1+\beta_2)}+\frac{V_{1,1}-V_{2,1}}{\beta_2(\beta_1+\beta_2)}\right]=0$$

De acuerdo con la geometría ilustrada en la figura 32.5, $\Delta x = 3$, $\Delta y = 2$, $\beta_1 = \beta_2 = \alpha_2 = 1$ y $\alpha_1 = 0.94281$. Sustituyendo estos valores se obtiene

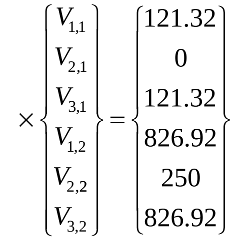
$$0.12132V_{1,1} - 121.32 + 0.11438V_{1,1} - 0.11438V_{2,1} + 0.25V_{1,1}$$
$$+ 0.25V_{1,1} - 0.25\,V_{1,2} = 0$$

Agrupando términos se tiene

$$0.73570V_{1,1} - 0.11438V_{2,1} - 0.25V_{1,2} = 121.32$$

Un procedimiento similar se aplica a los nodos interiores restantes. Las ecuaciones simultáneas resultantes se expresan en forma matricial como

$$\begin{bmatrix} 0.73570 & -0.11438 & & -0.25000 & & \\ -0.11111 & 0.72222 & -0.11111 & & -0.25000 & \\ & -0.11438 & 0.73570 & & & -0.25000 \\ -0.31288 & & & 1.28888 & -0.14907 & \\ & -0.25000 & & -0.11111 & 0.72222 & -0.11111 \\ & & -0.31288 & & -0.14907 & 1.28888 \end{bmatrix}$$



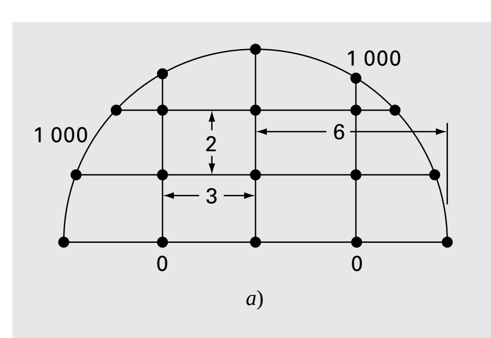
$$\times\begin{Bmatrix} V_{1,1} \\ V_{2,1} \\ V_{3,1} \\ V_{1,2} \\ V_{2,2} \\ V_{3,2} \end{Bmatrix}=\begin{Bmatrix} 121.32 \\ 0 \\ 121.32 \\ 826.92 \\ 250 \\ 826.92 \end{Bmatrix}$$

las cuales se resuelven para obtener

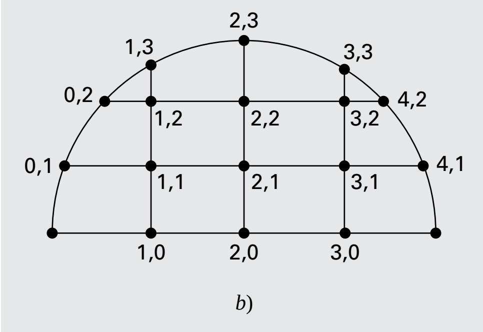
$$V_{1,1} = 521.19 \qquad V_{2,1} = 421.85 \qquad V_{3,1} = 521.19$$
$$V_{1,2} = 855.47 \qquad V_{2,2} = 755.40 \qquad V_{3,2} = 855.47$$

Estos resultados se representan en la figura 32.6*a*.

Para calcular el flujo (recuerde la sección 29.2.3), las ecuaciones (29.14) y (29.15) deben modificarse para tomar en cuenta las fronteras irregulares. En este ejemplo, las modificaciones dan por resultado



a)



b)

Figura 32.5
a) Sistema en dos dimensiones con un voltaje de 1 000 a lo largo de la frontera circular y un voltaje de 0 a lo largo de la base. *b*) Esquema de numeración nodal.

$$D_x = -\varepsilon \frac{V_{i+1,j} - V_{i-1,j}}{(\alpha_1 + \alpha_2)\Delta x}$$

y

$$D_y = -\varepsilon \frac{V_{i,j+1} - V_{i,j-1}}{(\beta_1 + \beta_2)\Delta y}$$

En el nodo (1, 1), estas fórmulas se utilizan para calcular las componentes x y y del flujo

$$D_x = -2 \frac{421.85 - 1\,000}{(0.94281 + 1)3} = 198.4$$

y

$$D_y = -2 \frac{855.47 - 0}{(1+1)2} = -427.7$$

las cuales se usan para calcular el vector de densidad de flujo eléctrico

$$D = \sqrt{198.4^2 + (-427.7)^2} = 471.5$$

con una dirección de

$$\theta = \tan^{-1}\left(\frac{-427.7}{198.4}\right) = -65.1°$$

Los resultados para los demás nodos son

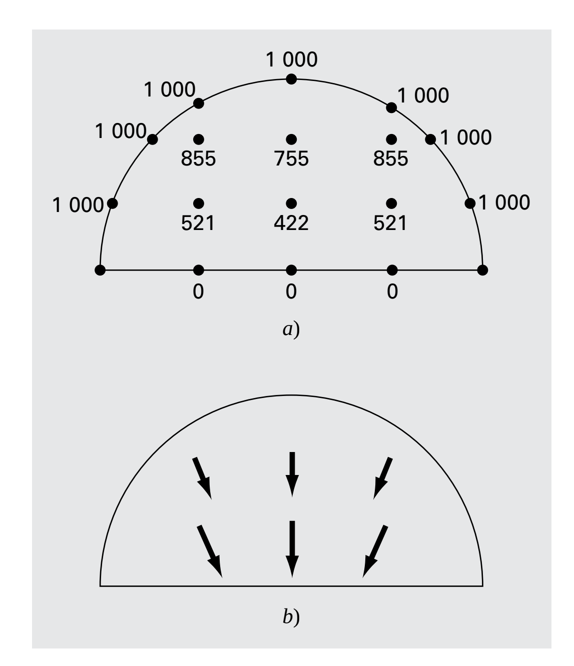


Figura 32.6
Resultados de la ecuación de Laplace con factores de corrección para las fronteras irregulares. *a*) Potencial y *b*) flujo.

Nodo	D_x	D_Y	D	θ
2, 1	0.0	–377.7	377.7	–90
3, 1	–198.4	–427.7	471.5	245.1
1, 2	109.4	–299.6	281.9	–69.1
2, 2	0.0	–289.1	289.1	–90.1
3, 2	–109.4	–299.6	318.6	249.9

Los flujos se muestran en la figura 32.6*b*.

32.4 SOLUCIÓN POR ELEMENTO FINITO DE UNA SERIE DE RESORTES (INGENIERÍA MECÁNICA/AERONÁUTICA)

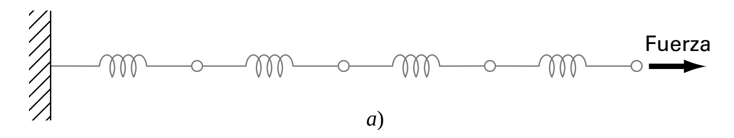
Antecedentes En la figura 32.7 se presenta una serie de resortes conectados entre sí. Un extremo está fijo a una pared; mientras que el otro está sujeto a una fuerza constante F. Usando el procedimiento paso por paso, descrito en el capítulo 31, se puede emplear un método por elemento finito para determinar los desplazamientos de los resortes.

Solución

Discretización. La forma de dividir este sistema es, obviamente, tratar cada resorte como un elemento. Así, el sistema consiste de cuatro elementos y cinco nodos (figura 32.7*b*).

Ecuaciones de los elementos. Como este sistema es muy simple, las ecuaciones de sus elementos se pueden dar directamente, sin recurrir a aproximaciones matemáticas. Éste es un ejemplo del procedimiento directo para deducir elementos.

En la figura 32.8 se muestra un solo elemento. La relación entre la fuerza F y el desplazamiento x se representa matemáticamente por la ley de Hooke:



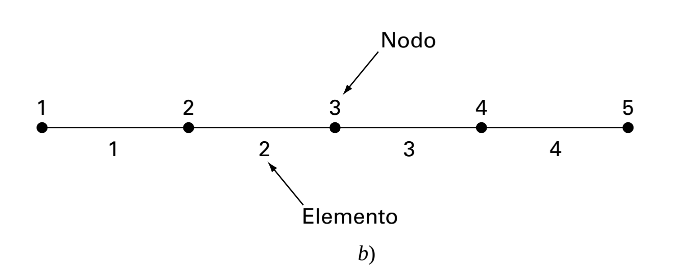


Figura 32.7
a) Serie de resortes conectados entre sí. Un extremo está fijo a una pared, mientras que el otro está sujeto a una fuerza constante *F*. *b*) Representación del elemento finito. Cada resorte representa un elemento. Por lo tanto, el sistema consiste de cuatro elementos y cinco nodos.

$$F = kx$$

donde k = la constante del resorte, que se interpreta como la fuerza requerida para causar un desplazamiento unitario. Si una fuerza F_1 se aplica al nodo 1, el siguiente balance de fuerzas debe satisfacerse:

$$F = k(x_1 - x_2)$$

donde x_1 = desplazamiento del nodo 1 desde su posición de equilibrio y x_2 = desplazamiento del nodo 2 desde su posición de equilibrio. Así, $x_2 - x_1$ representa cuánto se ha alargado o comprimido en relación con el equilibrio (figura 32.8).

Esta ecuación también se puede escribir como

$$F_1 = kx_1 - kx_2$$

Para un sistema en estado estacionario, un balance de fuerzas también necesita que $F_1 = -F_2$ y, por lo tanto,

$$F_2 = -kx_1 + kx_2$$

Estas dos ecuaciones simultáneas especifican el comportamiento del elemento en respuesta a las fuerzas dadas. Se escriben en forma matricial como

$$\begin{bmatrix} k & -k \\ -k & k \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} = \begin{Bmatrix} F_1 \\ F_2 \end{Bmatrix}$$

o

$$[k]\{x\} = \{F\} \tag{32.18}$$

donde la matriz $[k]$ es la matriz de las propiedades del elemento; en este caso, también se conoce como *matriz de rigidez del elemento*. Observe que la ecuación (32.18) se presenta en el formato de la ecuación (31.9). Así, se logró generar una ecuación matricial que describe el comportamiento de un elemento típico en nuestro sistema.

Antes de continuar con el siguiente paso (el ensamble de la solución total), presentaremos alguna notación. A los elementos de $[k]$ y $\{F\}$ se les colocan, de manera convencional, superíndices y subíndices:

$$\begin{bmatrix} k_{11}^{(e)} & -k_{12}^{(e)} \\ -k_{21}^{(e)} & k_{22}^{(e)} \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} = \begin{Bmatrix} F_1^{(e)} \\ F_2^{(e)} \end{Bmatrix}$$

donde el superíndice (e) indica que éstas son las ecuaciones del elemento. A las k también se les han puesto subíndices como k_{ij} para denotar su localización en el i-ésimo renglón y la j-ésima columna de la matriz. En este caso, también se interpretan físicamente como representación de la fuerza requerida en el nodo i para inducir un desplazamiento unitario en el nodo j.

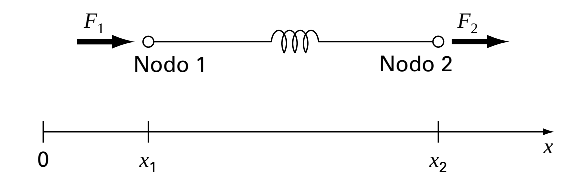


Figura 32.8
Diagrama de cuerpo libre de un sistema de resorte.

Ensamble. Antes de ensamblar las ecuaciones de los elementos, deben numerarse todos los elementos y nodos. Este esquema de numeración global especifica una configuración o topología del sistema. (Observe que en este caso se utiliza un esquema idéntico

al de la tabla 31.1.) Es decir, nos dice qué nodos pertenecen a qué elemento. Una vez que se especifica la topología, se pueden dar las ecuaciones de cada elemento con referencia a las coordenadas globales.

Las ecuaciones del elemento se agregan, una por una, para ensamblar todo el sistema. El resultado final se expresa en forma matricial como [recuerde la ecuación (31.10)]

$$[k]\{x'\} = \{F'\}$$

donde

$$[k]=\begin{bmatrix} k_{11}^{(1)} & -k_{12}^{(1)} & & & \\ -k_{21}^{(1)} & k_{22}^{(1)}+k_{11}^{(2)} & -k_{12}^{(2)} & & \\ & -k_{21}^{(2)} & k_{22}^{(2)}+k_{11}^{(3)} & -k_{12}^{(3)} & \\ & & -k_{21}^{(3)} & k_{22}^{(3)}+k_{11}^{(4)} & -k_{12}^{(4)} \\ & & & -k_{21}^{(4)} & k_{22}^{(4)} \end{bmatrix} \tag{32.19}$$

y

$$[F']=\begin{Bmatrix} F_1^{(1)} \\ 0 \\ 0 \\ 0 \\ F_2^{(4)} \end{Bmatrix}$$

y $\{x'\}$ y $\{F'\}$ son los desplazamientos expandidos y vectores de fuerza, respectivamente. Observe que, conforme las ecuaciones fueron ensambladas, las fuerzas internas se cancelan. Así, en el resultado final $\{F'\}$ tiene cero en todos los nodos, excepto en el primero y en el último.

Antes de continuar con el siguiente paso, debemos hacer una observación sobre la estructura de la matriz de propiedades de ensamble [ecuación (32.19)]. Observe que la matriz es tridiagonal, que es un resultado directo del esquema de numeración global particular que se eligió (tabla 31.1) antes del ensamblado. Aunque no es muy importante en el presente contexto, la obtención de tal sistema disperso en banda puede ser una ventaja decisiva en problemas más complicados. Ello se debe a los eficientes esquemas para resolver tales sistemas.

Condiciones de frontera. El presente sistema está sujeto a una sola condición de frontera, $x_1 = 0$. La introducción de esta condición y la aplicación del esquema de renumeración global reduce el sistema a (todas las $k = 1$)

$$\begin{bmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 1 \end{bmatrix} \begin{Bmatrix} x_2 \\ x_3 \\ x_4 \\ x_5 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ F \end{Bmatrix}$$

El sistema ahora tiene la forma de la ecuación (31.11) y está listo para resolverse.

Aunque la reducción de las ecuaciones es ciertamente un método correcto para incorporar condiciones de frontera, por lo común se prefiere dejar intacto el número de ecuaciones cuando se obtiene la solución en la computadora. Sea cual fuere el método, una vez que se incorporan las condiciones de frontera, es posible llegar al paso siguiente: la solución.

Generación de la solución. Usando uno de los procedimientos de la parte tres, tal como la eficiente técnica de solución tridiagonal descrita en el capítulo 11, el sistema se resuelve para obtener (con todas las $k = 1$ y $F = 1$)

$$x_2 = 1 \qquad x_3 = 2 \qquad x_4 = 3 \qquad x_5 = 4$$

Procesamiento posterior. Los resultados pueden mostrarse ahora en forma gráfica. Como en la figura 32.9, los resultados son los que se esperaban. Cada resorte se estira un desplazamiento unitario.

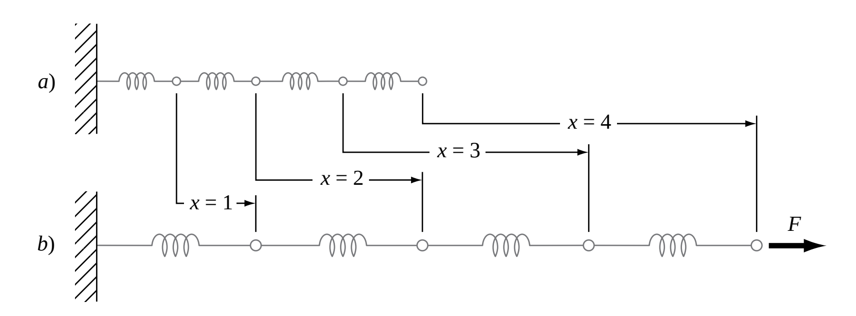


Figura 32.9
a) El sistema de resortes original. *b*) El sistema después de la aplicación de una fuerza constante. Los desplazamientos se indican en el espacio entre los dos sistemas.

PROBLEMAS

Ingeniería química/bioingeniería

32.1 Realice el mismo cálculo de la sección 32.1, pero ahora use $\Delta x = 1.25$.

32.2 Desarrolle una solución por elemento finito para el sistema en estado estacionario de la sección 32.1.

32.3 Calcule los flujos de masa para la solución en estado estacionario de la sección 32.1 usando la primera ley de Fick.

32.4 Calcule la distribución en estado estacionario de la concentración en el reactor mostrado en la figura P32.4. La EDP que rige este sistema es

$$D\left(\frac{\partial^2 c}{\partial x^2}+\frac{\partial^2 c}{\partial y^2}\right)-kc=0$$

y las condiciones de frontera son las que se muestran. Emplee un valor de 0.5 para D y 0.1 para k.

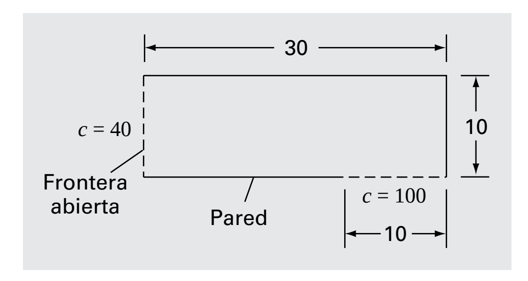


Figura P32.4

32.5 Entre dos placas hay una separación de 10 cm, como se muestra en la figura P32.5. Inicialmente, ambas placas y el fluido están en reposo. En $t = 0$, la placa superior se mueve con una velocidad constante de 8 cm/s. Las ecuaciones que rigen los movimientos de los fluidos son

$$\frac{\partial v_{\text{aceite}}}{\partial t}=\mu_{\text{aceite}}\frac{\partial^2 v_{\text{aceite}}}{\partial x^2} \quad \text{y} \quad \frac{\partial v_{\text{agua}}}{\partial t}=\mu_{\text{agua}}\frac{\partial^2 v_{\text{agua}}}{\partial x^2}$$

y las siguientes relaciones se satisfacen en la interfaz aceite-agua

$$v_{\text{aceite}}=v_{\text{agua}} \quad \text{y} \quad \mu_{\text{aceite}}\frac{\partial v_{\text{aceite}}}{\partial x}=\mu_{\text{agua}}\frac{\partial v_{\text{agua}}}{\partial x}$$

¿Cuál es la velocidad de las dos capas de fluido en $t = 0.5$, 1 y 1.5 s, a las distancias $x = 2$, 4, 6 y 8 cm de la placa inferior? Observe que μ_{agua} y μ_{aceite} = 1 y 3 cp, respectivamente.

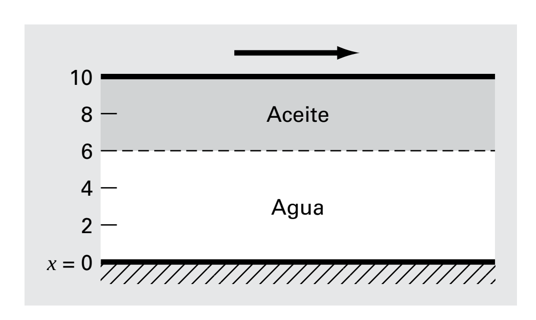


Figura P32.5

32.6 El desplazamiento de una membrana uniforme sujeta a una tensión y una presión uniforme se puede describir por la ecuación de Poisson

$$\frac{\partial^2 z}{\partial x^2}+\frac{\partial^2 z}{\partial y^2}=-\frac{P}{T}$$

Despeje el desplazamiento de una membrana cuadrada de 1 cm que tiene $P/T = 0.6$/cm y está fijada de modo que tiene desplazamiento cero a lo largo de sus cuatro fronteras. Utilice $\Delta x = \Delta y = 0.1$ cm. Muestre sus resultados como una gráfica de contorno.

Ingeniería civil/ambiental

32.7 Ejecute el mismo cálculo que en la sección 32.2, pero utilice $\Delta x = \Delta y = 0.4$ m.

32.8 El flujo a través de medios porosos queda descrito con la ecuación de Laplace

$$\frac{\partial^2 h}{\partial x^2}+\frac{\partial^2 h}{\partial y^2}=0$$

donde h es la carga. Use métodos numéricos para determinar la distribución de la carga para el sistema que se muestra en la figura P32.8.

32.9 La velocidad del flujo del agua a través de los medios porosos se relaciona con la carga por medio de la ley de D'Arcy

$$q_n=-K\frac{dh}{dn}$$

donde K es la conductividad hidráulica y q_n es la velocidad de descarga en la dirección n. Si $K = 5 \times 10^{-4}$ cm/s, calcule la velocidad del agua para el problema 32.8.

Ingeniería eléctrica

32.10 Realice el mismo cálculo que en la sección 32.3, pero para el sistema que se ilustra en la figura P32.10.

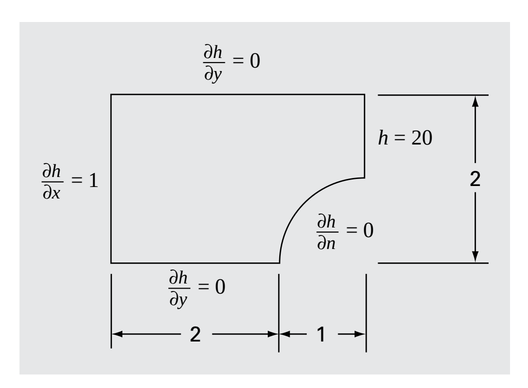


Figura P32.8

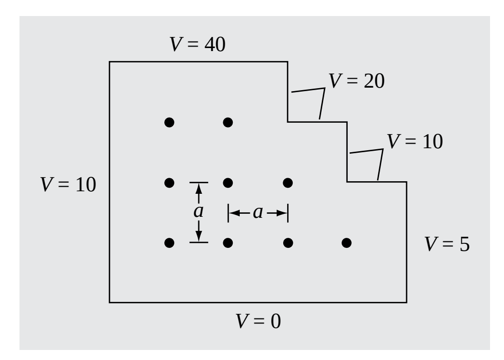


Figura P32.10

32.11 Lleve a cabo el mismo cálculo que en la sección 32.3, pero para el sistema que se muestra en la figura P32.11.

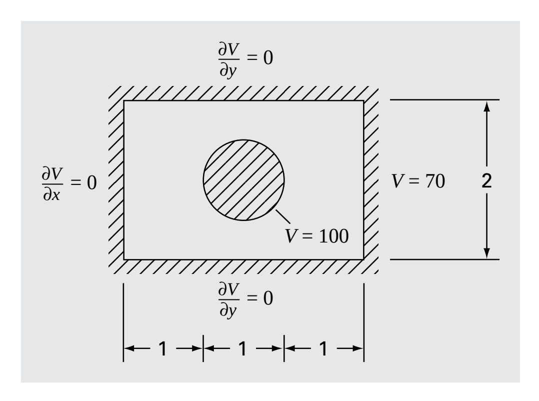


Figura P32.11

32.12 Use la ecuación de Poisson para calcular el potencial eléctrico sobre una placa cuadrada unitaria (1 × 1) con voltaje cero en las orillas y fuentes puntuales de carga de $\rho_v/\varepsilon(0.7, 0.7) = 1$ y $\rho_v/\varepsilon(0.3, 0.3) = -1$. Utilice $\Delta x = \Delta y = 0.1$ y muestre sus resultados como gráfica de contorno.

Ingeniería mecánica/aeroespacial

32.13 Lleve a cabo el mismo cálculo que en la sección 32.4, pero cambie la fuerza a 1.5 y las constantes de los resortes a

Resorte	1	2	3	4
k	0.75	1.5	0.5	2

32.14 Efectúe el mismo cálculo que en la sección 32.4, pero utilice una fuerza de 2 y cinco resortes con

Resorte	1	2	3	4	5
k	0.25	0.5	1.5	0.75	1

32.15 Una barra compuesta y aislada está formada por dos partes sujetas extremo con extremo, ambas con la misma longitud. La parte a tiene conductividad térmica k_a, para $0 \le x \le ½$, y la parte b tiene conductividad térmica k_b, para $½ \le x \le 1$. Las ecuaciones de conducción de calor transitivas no dimensionales que describen la temperatura u en la longitud x de la barra compuesta son

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t} \qquad 0 \le x \le 1/2$$

$$r\frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t} \qquad 1/2 \le x \le 1$$

donde u = temperatura, x = coordenada axial, t = tiempo y $r = k_a/k_b$. Las condiciones iniciales y de frontera son

Condiciones de frontera	$u(0, t) = 1$	$u(1, t) = 1$
	$\left(\frac{\partial u}{\partial x}\right)_a = \left(\frac{\partial u}{\partial x}\right)_b$	$x = 1/2$
Condiciones iniciales	$u(x, 0) = 0$	$0 < x < 1$

Resuelva este conjunto de ecuaciones para la distribución de temperatura como función del tiempo. Utilice análogos de diferencias finitas exactas de segundo orden para las derivadas, con formulación de Crank-Nicolson, para integrar en el tiempo. Escriba un programa de computadora para la solución, y seleccione valores de Δx y Δt para una buena exactitud. Grafique la temperatura u *versus* la longitud x para distintos valores de tiempo t. Genere una curva separada para los valores siguientes del parámetro r = 1, 0.1, 0.01, 0.001 y 0.

32.16 Resuelva la ecuación de conducción del calor transitiva no dimensional en dos dimensiones, que representa la distribución de temperatura transitiva en una placa aislada. La ecuación gobernante es

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{\partial u}{\partial t}$$

donde u = temperatura, x y y son las coordenadas espaciales y t = tiempo. Las condiciones iniciales y de frontera son

Condiciones de frontera	$u(x, 0, t) = 0$	$u(x, 1, t) = 1$	
	$u(0, y, t) = 0$	$u(1, y, t) = 1$	
Condiciones iniciales	$u(x, y, 0) = 0$	$0 \le x < 1$	$0 \le y < 1$

Resuelva con el empleo de una técnica alternativa de dirección implícita. Escriba un programa de cómputo para implementar la solución. Grafique los resultados con el uso de una rutina graficadora de tres dimensiones en la que el plano horizontal contenga los ejes x y y, y el eje z es la variable dependiente u. Haga varias gráficas en distintos tiempos, incluyendo lo siguiente *a*) las condiciones iniciales; *b*) un tiempo intermedio, aproximadamente a la mitad del camino hacia el estado estable y *c*) la condición de estado estable.

EPÍLOGO: PARTE OCHO

PT8.3 ALTERNATIVAS

Las principales ventajas y desventajas asociadas a los métodos numéricos para la solución de ecuaciones diferenciales parciales implican seleccionar entre procedimientos por *diferencias finitas* y por *elemento finito*. Los métodos por diferencias finitas son conceptualmente más fáciles de comprender. Además, son de fácil programación en sistemas que pueden ser aproximados con mallas uniformes. Sin embargo, son difíciles de aplicar a sistemas con geometrías complicadas.

Los procedimientos por diferencias finitas se dividen en categorías, dependiendo del tipo de EDP que se vaya a resolver. Las *EDP elípticas* pueden aproximarse por medio de un conjunto de ecuaciones algebraicas lineales. En consecuencia, el *método de Liebmann* (que, de hecho, es el método de Gauss-Seidel) se utiliza para obtener una solución de manera iterativa.

Las *EDP parabólicas en una dimensión* se resuelven de dos maneras fundamentalmente diferentes: con métodos explícitos o con métodos implícitos. El *método explícito* se desarrolla en el tiempo de una forma similar a la técnica de Euler para resolver las EDO. Tiene la ventaja de que se programa fácilmente, aunque presenta el inconveniente de tener un criterio de estabilidad muy estricto. En cambio, existen métodos implícitos que, por lo general, implican la solución de ecuaciones algebraicas tridiagonales de manera simultánea en cada iteración. Uno de esos procedimientos, el *método de Crank-Nicolson*, es muy preciso y estable, y, por lo tanto, es muy utilizado en problemas parabólicos lineales en una dimensión.

Las *EDP parabólicas en dos dimensiones* también se modelan de manera explícita. Aunque sus restricciones de estabilidad son aún más estrictas que en el caso de una dimensión. Se han desarrollado procedimientos implícitos especiales (generalmente conocidos como métodos de separación) para evitar dicho inconveniente. Estos procedimientos son eficientes y estables. Uno de los más comunes es el método *implícito de dirección alternante* o *IDA*.

Todos los procedimientos por *diferencias finitas* anteriores se vuelven complicados cuando se aplican a sistemas con formas no uniformes y condiciones heterogéneas. Existen métodos por elemento finito que funcionan mejor para tales sistemas.

Aunque el *método del elemento finito* se basa en ideas muy sencillas, el mecanismo para generar un buen código del elemento finito para problemas en dos y tres dimensiones no es un ejercicio trivial. Además, llega a ser costoso en términos computacionales para problemas grandes. Sin embargo, es muy superior a los procedimientos por diferencias finitas para sistemas con formas complicadas. En consecuencia, a menudo se justifica su costo debido a su concepto “superior” en el detalle de la solución final.

PT8.4 RELACIONES Y FÓRMULAS IMPORTANTES

En la tabla PT8.3 se resume la información importante que fue presentada en la parte ocho respecto de los métodos por diferencias finitas. Esta tabla es útil para lograr un rápido acceso a relaciones y fórmulas importantes.

PT8.5 MÉTODOS AVANZADOS Y REFERENCIAS ADICIONALES

Carnahan, Luther y Wilkes (1969); Rice (1983); Ferziger (1981), y Lapidus y Pinder (1982) ofrecen análisis útiles de los métodos y del software para resolver EDP. También se recomienda consultar Ames (1977), Gladwell y Wait (1979), Vichnevetsky (1981, 1982) y Zienkiewicz (1971) para estudios más profundos. Existe información adicional sobre el método del elemento finito en Allaire (1985), Huebner y Thornton (1982), Stasa (1985) y Baker (1983). Además de las EDP elípticas e hiperbólicas, también existen métodos numéricos para resolver EDP hiperbólicas. Se encuentran buenas introducciones y resúmenes de algunos de tales métodos en Lapidus y Pinder (1981), Ferziger (1981), Forsythe y Wasow (1960) y Hoffman (1992).

Tabla PT8.3 Resumen de los métodos por diferencias finitas.

	Molécula computacional	Ecuación
EDP elípticas Método de Liebmann	$i, j+1$ $i-1, j$ i, j $i+1, j$ $i, j-1$	$T_{i,j} = \dfrac{T_{i+1,j} + T_{i-1,j} + T_{i,j+1} + T_{i,j-1}}{4}$
EDP parabólicas (en una dimensión) Método explícito	$i, l+1$ $i-1, l$ i, l $i+1, l$	$T_i^{l+1} = T_i^l + \lambda(T_{i+1}^l - 2T_i^l + T_{i-1}^l)$
Método implícito	$i-1, l+1$ $i, l+1$ $i+1, l+1$ i, l	$-\lambda T_{i-1}^{l+1} + (1+2\lambda)T_i^{l+1} - \lambda T_{i+1}^{l+1} = T_i^l$
Método de Crank-Nicolson	$i-1, l+1$ $i, l+1$ $i+1, l+1$ $i, l+\frac{1}{2}$ $i-1, l$ i, l $i+1, l$	$-\lambda T_{i-1}^{l+1} + 2(1+\lambda)T_i^{l+1} - \lambda T_{i+1}^{l+1}$ $= \lambda T_{i-1}^l + 2(1-\lambda)T_i^l + \lambda T_{i+1}^l$

APÉNDICE A

La serie de Fourier

La serie de Fourier puede expresarse de diferentes maneras. Dos expresiones trigonométricas equivalentes son

$$f(t) = a_0 + \sum_{k=1}^{\infty} [a_k \cos(k\omega_0 t) + b_k \operatorname{sen}(k\omega_0 t)]$$

o

$$f(t) = a_0 + \sum_{k=1}^{\infty} [c_k \cos(k\omega_0 t + \theta_k)]$$

donde los coeficientes están relacionados mediante (véase figura A.1)

$$c_k = \sqrt{a_k^2 + b_k^2}$$

y

$$\theta_k = -\tan^{-1}\left(\frac{b_k}{a_k}\right)$$

Además de las formas trigonométricas, las series también se expresan en términos de la función exponencial,

$$f(t) = \tilde{c}_0 + \sum_{k=1}^{\infty} [\tilde{c}_k e^{ik\omega_0 t} + \tilde{c}_{-k} e^{-ik\omega_0 t}] \tag{A.1}$$

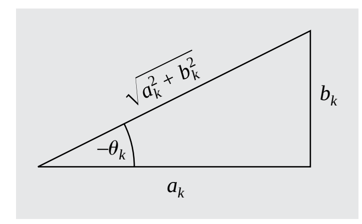


Figura A.1
Relaciones entre las formas rectangular y polar de los coeficientes de la serie de Fourier.

donde (véase figura A.2)

$$\tilde{c}_0 = a_0$$

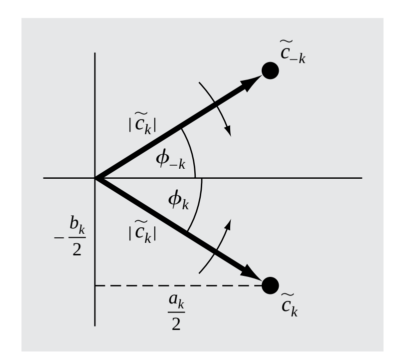


Figura A.2
Relaciones entre coeficientes exponenciales complejos y reales de la serie de Fourier.

$$\tilde{c}_k = \frac{1}{2}(a_k - ib_k) = |\tilde{c}_k| e^{i\phi} k$$

$$\tilde{c}_{-k} = \frac{1}{2}(a_k + ib_k) = |\tilde{c}_k| e^{-i\phi} k$$

donde $|\tilde{c}_0| = a_0$ y

$$|\tilde{c}_k| = \frac{1}{2}\sqrt{a_k^2 + b_k^2} = \frac{c_k}{2}$$

y

$$\phi_k = \tan^{-1}\left(\frac{-b_k}{a_k}\right)$$

Observe que la tilde significa que el coeficiente es un número complejo.

Cada término de la ecuación (A.1) puede visualizarse como un fasor giratorio (las flechas de la figura A.2). Los términos con subíndice positivo giran en sentido contrario a las manecillas de un reloj analógico; mientras que los que tienen subíndice negativo giran en sentido de las manecillas del reloj. Los coeficientes $\tilde{c}_k$ y $\tilde{c}_{-k}$ especifican la posición del fasor en $t = 0$. Entonces, la suma infinita de los fasores, que se dejan girar en $t = 0$, es igual a $f(t)$.

APÉNDICE B

Empecemos con MATLAB

MATLAB es un programa computacional que ofrece al usuario un ambiente adecuado para diversos tipos de cálculos (en particular, aquellos que se relacionan con manipulaciones de matrices). MATLAB opera interactivamente ejecutando, una por una, las instrucciones del usuario, conforme se van introduciendo. Se puede guardar una serie de instrucciones como un guión y correrlas como un programa interpretativo. MATLAB tiene muchas funciones predefinidas; sin embargo, es posible que los usuarios construyan sus propias funciones a partir de comandos y funciones de MATLAB. Las principales características de MATLAB son cálculos predefinidos con vectores y matrices como:

- Aritmética de vectores y matrices
- Inversión de matrices y análisis de valores y vectores propios
- Aritmética compleja y operaciones con polinomios
- Cálculos estadísticos
- Despliegue de gráficas
- Diseño de sistemas de control
- Modelos de proceso de ajuste a partir del análisis de datos

MATLAB tiene diferentes paqueterías que proporcionan funciones especializadas. Éstas incluyen procesamiento de señales, sistemas de control, identificación de sistemas, optimización y estadística.

MATLAB está disponible en versiones para PC, Mac y estaciones de trabajo. La versión moderna para PC opera en ambiente Windows. Los siete ejercicios siguientes están diseñados para que puedan ser calculados con MATLAB; no constituyen un tutorial completo. Existen materiales tutoriales adicionales en los manuales de MATLAB. Un gran número de libros de texto ofrecen ejercicios con MATLAB. También se dispone de información en línea para cualquier comando o función, tecleando `help` *`name`*, donde *`name`* identifica el comando. No se limite solamente a estos ejercicios; aparte de probarlos todos, intente las variaciones que se le puedan ocurrir. Compruebe las respuestas que da MATLAB, asegurándose de entenderlas y de que sean correctas. Ésta es la manera más efectiva de aprender MATLAB.

1. Asignación de valores a los nombres de las variables

La asignación de valores a variables escalares es similar a otros lenguajes de computación. Teclee

```
a = 4
```

y

```
A = 6
```

Observe cómo la asignación se repite para confirmar lo que usted ha hecho. Ésta es una característica de MATLAB. La repetición se elimina terminando la línea de instrucción con un punto y coma (`;`). Teclee

```
b = -3;
```

MATLAB considera a los nombres reconociendo mayúsculas y minúsculas; es decir, el nombre `a` no es lo mismo que el nombre `A`. Para ilustrar esto, introduzca

```
a
```

y

```
A
```

Vea cómo sus valores son distintos. Son nombres distintos.

En MATLAB, los nombres de las variables generalmente representan cantidades matriciales. Un vector renglón se puede asignar como sigue:

```
a = [ 1 2 3 4 5 ]
```

Nuevamente la repetición confirma la asignación. Advierta cómo se ha tomado la nueva asignación de a. Un vector columna se introduce de varias maneras. Pruébelo.

```
b = [ 1 ; 2 ; 3 ; 4 ; 5 ]
```

o

```
b = [ 1;
      2;
      3;
      4;
      5 ]
```

o, transponiendo un vector renglón con el operador ′

```
b = [ 1 2 3 4 5 ]′
```

Una matriz de valores en dos dimensiones se asigna como sigue:

```
A = [ 1 2 3 ; 4 5 6 ; 7 8 8 ]
```

o

```
A = [ 1 2 3 ;
      4 5 6 ;
      7 8 8  ]
```

Los valores almacenados por una variable pueden examinarse en cualquier momento tecleando tan sólo el nombre; por ejemplo:

```
b
```

o

```
A
```

Además, se obtiene una lista de todas las variables en uso con la instrucción

```
who
```

o si se quiere con más detalle introduzca

```
whos
```

Existen algunas variables predefinidas; por ejemplo, pi.

También se pueden asignar valores complejos a las variables, ya que MATLAB trata automáticamente la aritmética compleja. Para hacerlo, es conveniente asignar un nombre de variable, por lo general i o j, para la raíz cuadrada de –1.

```
i = sqrt(-1)
```

Después se asigna un valor complejo así

```
x = 2 + i*4
```

2. Operaciones matemáticas

Las operaciones con cantidades escalares se llevan a cabo de manera directa, como en otros lenguajes de computación. Los operadores comunes, en orden de prioridad, son

```
^     Exponenciación
* /   Multiplicación y división
\     División por la izquierda (se aplica a matrices)
+ -   Adición y sustracción
```

Estos operadores funcionan como en las calculadoras. Hagamos

```
2 * pi
```

También se pueden incluir variables escalares reales:

```
y = pi / 4
y ^ 2.45
```

Los resultados de los cálculos pueden ser asignados a una variable (como en los dos penúltimos ejemplos) o tan sólo se despliegan (como en el último ejemplo).

Los cálculos también pueden utilizar cantidades complejas. Usando la `x` definida anteriormente,

```
3 * x
1 / x
x ^ 2
x + y
```

La verdadera potencia de MATLAB se muestra en su capacidad para realizar cálculos con matrices. El producto interno de dos vectores (producto punto) se calcula usando el operador `*`

```
a * b
```

y, de la misma forma, el producto externo

```
b * a
```

Para ilustrar la multiplicación de un vector por una matriz, primero redefinimos `a` y `b`,

```
a = [ 1 2 3 ]
```

y

```
b = [ 4 5 6 ]'
```

Ahora, hagamos

```
a * A
```

o

```
A * b
```

¿Qué ocurre cuando las dimensiones no son las requeridas por las operaciones? Para verlo, escribamos

```
A * a
```

La multiplicación de una matriz por otra matriz se lleva a cabo de forma similar:

```
A * A
```

También se pueden hacer operaciones con escalares:

```
A / pi
```

Es importante recordar que MATLAB aplicará las operaciones aritméticas simples en forma de vector y matriz, si es posible. En ocasiones, usted necesitará realizar los cálculos elemento por elemento en una matriz o vector. MATLAB también puede hacer esto. Por ejemplo,

```
A ^ 2
```

da como resultado una multiplicación matricial de `A` consigo misma. Pero, ¿qué hacer si queremos elevar al cuadrado cada elemento de `A`? Esto se efectúa con

```
A . ^ 2
```

El . que precede al operador `^` significa que la operación será llevada a cabo elemento por elemento. En el manual de MATLAB se le llama *operaciones de arreglos.*

Cuando se utiliza el operador división (`/`) con matrices, el uso de una matriz inversa está implícito. Por lo tanto, si `A` es una matriz cuadrada no singular, entonces `B/A` corresponde a la multiplicación por la derecha de `B` por la inversa de `A`. Un camino largo para hacerlo consiste en usar la

función *inv*; es decir, B*inv(A); sin embargo, usar el operador división es más eficiente ya que X = B/A en realidad resuelve el conjunto de ecuaciones X*A=B usando un esquema de descomposición/eliminación.

La "división por la izquierda" (\ , diagonal invertida) se emplea también en las operaciones con matrices. Así, A\B corresponde a la multiplicación por la izquierda de B por la inversa de A. De esta manera se resuelve el conjunto de ecuaciones A*X=B, un cálculo común en ingeniería.

Por ejemplo si c es un vector columna con valores 0.1, 1.0 y 10, la solución de A * x = c, donde A fue definida antes, se obtiene al escribir

```
c = [ 0.1 1.0 10 ]'
x = A \ c
```

Inténtelo.

3. Uso de funciones predefinidas

MATLAB y sus paqueterías tienen una amplia colección de funciones predefinidas. Usted puede usar la ayuda en línea para encontrar más información acerca de ellas. Una de sus propiedades importantes es que operan directamente sobre cantidades vectoriales y matriciales. Por ejemplo, intente

```
log(A)
```

y verá que la función logaritmo natural se aplica en un estilo de arreglo, elemento por elemento, a la matriz A. La mayoría de las funciones, como *sqrt, abs, sin, acos, tanh, exp*, operan en forma de arreglo. También ciertas funciones, como la exponencial y la raíz cuadrada, tienen definiciones de matriz. MATLAB evaluará la versión matricial cuando se agregue la letra m al nombre de la función. Intente

```
sqrtm(A)
```

Un uso común de las funciones consiste en evaluar una fórmula para una serie de argumentos. Construya un vector columna t que contenga valores desde 0 hasta 100 a intervalos de 5,

```
t = [ 0 : 5 : 100 ]'
```

Compruebe el número de entradas en el arreglo t con la función *Length*,

```
length(t)
```

Ahora, supongamos que quiere evaluar una fórmula y = f(t), donde la fórmula se calcula para cada valor del arreglo de t, y el resultado se asigna a una posición correspondiente en el arreglo y. Por ejemplo,

```
y = t .^ 0.34 – log10(t) + 1 ./ t
```

¡Listo! [Observe el uso de los operadores del arreglo adyacentes a los puntos decimales.] Esto es similar a crear una columna con los valores t en una hoja de cálculo, y copiar una fórmula en una columna adyacente para evaluar los valores de y.

4. Gráficas

Las capacidades gráficas de MATLAB son similares a las de un programa en una hoja de cálculo. Las gráficas se crean rápida y de manera conveniente; sin embargo, no hay mucha flexibilidad para personalizarlas.

Por ejemplo, para crear una gráfica de los arreglos t, y anteriores, introduzca

```
plot(t, y)
```

¡Eso es! Ahora usted puede personalizar un poco la gráfica utilizando los siguientes comandos:

```
title('Grafica de y contra t')
xlabel('valores de t')
ylabel('valores de y')
grid
```

La gráfica aparece en otra ventana y puede imprimirse o transferirse a través del portapapeles (PC con Windows o Mac) a otros programas.

Existen otras características de las gráficas que serán de utilidad: trazo de gráficas de objetos en lugar de líneas, familias de curvas, trazo de gráficas en el plano complejo, ventanas de gráficas múltiples, gráficas log-log o semilog, gráficas tridimensionales y gráficas de contorno.

5. Polinomios

Existen muchas funciones de MATLAB que le permiten operar sobre los arreglos como si sus entradas fueran coeficientes o raíces de ecuaciones polinomiales. Por ejemplo, introduzca

```
c = [ 1 1 1 1 ]
```

y después

```
r = roots(c)
```

y las raíces del polinomio $x^3 + x^2 + x + 1 = 0$ se imprimirán y además se almacenarán en el arreglo `r`. Los coeficientes de un polinomio pueden calcularse a partir de las raíces con la función *poly*,

```
poly(r)
```

y un polinomio puede evaluarse para un valor dado de `x`. Por ejemplo,

```
polyval(c, 1.32)
```

Si otro polinomio, $2x^2 - 0.4x - 1$, se representa por el arreglo `d`,

```
d = [ 2 -0.4 -1 ]
```

los dos polinomios pueden multiplicarse simbólicamente con la función convolución, *conv*; para obtener los coeficientes del producto polinomial, escriba

```
cd = conv(c,d)
```

La función de deconvolución, *deconv*, sirve para dividir un polinomio entre otro; por ejemplo,

```
[ q,r ] = deconv(c,d)
```

El resultado `q` es el cociente, y el resultado `r` es el residuo.

Existen otras funciones polinomiales que son de utilidad, como la función residuo, que da la expansión en fracciones parciales.

6. Análisis estadístico

Los paquetes estadísticos contienen muchas características para el análisis estadístico; sin embargo, los cálculos estadísticos comunes se realizan con el conjunto básico de funciones de MATLAB. Usted puede generar una serie de números (seudo) aleatorios con la función *rand*. Se dispone de una distribución uniforme o normal:

```
n = 0 : 5 : 1000 ;
```

(¡Olvidó el `;` !)

```
num = rand(size(n)) ;
```

Probablemente entendió por qué es importante usar punto y coma al final de las instrucciones anteriores, en especial si no tuvo cuidado de usarlo.

Si desea ver una gráfica de ruido intente

```
plot(num)
```

Se supone que éstos son números distribuidos normalmente con una media de cero y una varianza (y desviación estándar) de uno. Compruébelo mediante

```
mean(num)
```

y

```
std(num)
```

¡Nadie es perfecto! Usted puede hallar los valores mínimos y máximos,

```
min(num)
max(num)
```

Hay una función adecuada para trazar un histograma de los datos:

```
hist(num,20)
```

donde 20 es el número de compartimientos.

Si quiere ajustar un polinomio para algunos datos por mínimos cuadrados, utilice la función *polyfit*. Intente el siguiente ejemplo:

```
t = 05
y = [ -0.45 0.56 2.34 5.6 9.45 24.59 ]
coef = polyfit( t, y, 3 )
```

Los valores de `coef` son los coeficientes del polinomio ajustados. Para generar el valor calculado de `y`,

```
yc = polyval( coef,t )
```

y para graficar los datos contra la curva ajustada,

```
plot ( t,yc,t,y,'o' )
```

La gráfica de la curva continua es lineal por partes; por lo tanto, no parece muy suave. Mejórela, así:

```
t1 = [ 0 : 0.05 : 5 ] ;
yc = polyval(coef, t1)
plot(t1, yc, t, y, 'o')
```

7. Esto y aquello

Existen muchas otras características de MATLAB. Algunas de ellas las encontrará útiles; otras quizá nunca las use. Le sugerimos que explore y experimente.

Para guardar una copia de su sesión, MATLAB tiene una posibilidad útil llamada *diary*. Utilice la instrucción

```
diary problem1
```

y MATLAB abre un archivo de disco donde almacena todas las instrucciones y los resultados (no las gráficas) de su sesión. Usted puede cerrar la instrucción *diary* tecleando:

```
diary off
```

y regresar al mismo archivo:

```
diary on
```

Después de salir de MATLAB, el archivo *diary* estará disponible. Es común usar un editor o procesador de palabras para limpiar el archivo *diary* (eliminando todos los errores que haya cometido, ¡antes de que otras personas los vean!) y después imprimir el archivo para obtener una copia de las partes importantes de su sesión de trabajo; por ejemplo, los resultados numéricos clave.

Salga de MATLAB con los comandos `quit` o `exit`. Se puede guardar el estado actual de su trabajo con la instrucción `save`. También se puede volver a cargar dicho estado con la instrucción `load`.

APÉNDICE C

Iniciación a Mathcad

Mathcad tiene una forma única de manejar las ecuaciones, los números, el texto y las gráficas. Funciona como un cuaderno de notas y un lápiz. La interfaz de pantalla es una hoja de cálculo en blanco en la que usted introduce ecuaciones, datos de gráficas o funciones, y anota las operaciones con texto. Cuando es posible, Mathcad usa símbolos matemáticos convencionales para representar los operadores matemáticos. Por lo tanto, usted puede descubrir que la interfaz de Mathcad es bastante natural y familiar.

Mathcad puede resolver una amplia gama de problemas matemáticos, ya sea en forma numérica o simbólica. Las capacidades simbólicas de Mathcad tienen relativamente poca aplicación en este texto, aunque se pueden usar para verificar nuestros resultados numéricos. Por lo tanto, no se cubrirán en detalle en esta descripción general. Mathcad tiene un conjunto amplio de operadores y funciones que dejan que usted realice mucho de los métodos numéricos que cubre este texto. También tiene un lenguaje de programación que le permite escribir sus propios procedimientos multilineales y subprogramas. La siguiente exposición presenta una breve descripción de las características de Mathcad que usted encontrará de máxima utilidad para este texto.

FUNDAMENTOS DE MATHCAD

En este texto, las aplicaciones exigirán que usted pueda crear sus propias hojas de cálculo. Para facilitar sus esfuerzos, se revisan a continuación las principales características de la ventana de aplicación Mathcad.

El menú principal



Ésta es la puerta de entrada a Mathcad. También proporciona los comandos que manejan los detalles de edición y administración de las hojas de cálculo. Por ejemplo, haga clic en los menús **File** y **Tools** para ver algo de la funcionalidad disponible para usted.

La barra estándar de herramientas



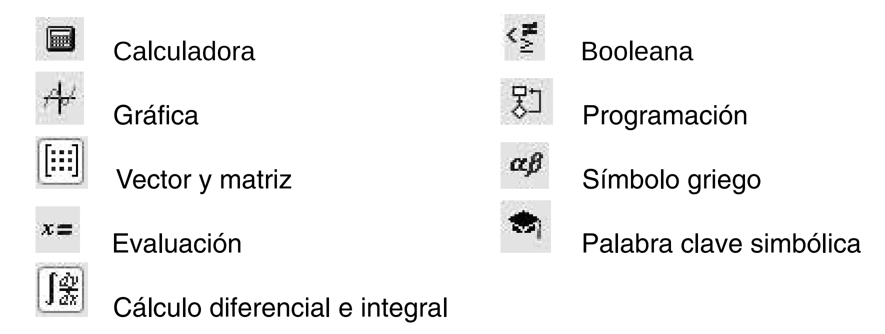
Deben aparecer automáticamente varias barras de herramienta debajo del menú **Main**. Como el nombre lo sugiere, la barra de herramientas **Standard** ofrece atajos para muchas de las tareas comunes, desde abrir una hoja de cálculo y guardar el expediente hasta mostrar listas de funciones y unidades interconstruidas.

Dependiendo de lo que usted esté haciendo en su hoja de cálculo, pueden aparecer desactivados en gris uno o más de estos botones. Si usted pasa el ratón sobre cada uno de los botones de la paleta, verá una descripción de la función del botón.

La Pallette Calculator



Puede ser que la **Pallette Math** aparezca automáticamente en la parte superior de la pantalla. En caso contrario, sólo seleccione **View**, **Toolbars**, **Math** y aparecerá. A continuación se describen los botones y sus funciones:



Haga clic en alguno de estos botones para hacer aparecer la pallette completa. Usted puede usar las paletas para insertar símbolos matemáticos y operaciones matemáticas directamente en su hoja de cálculo Mathcad.

INTRODUCCIÓN DE TEXTO Y OPERACIONES MATEMÁTICAS

Introducción de texto

Para crear una región de texto, haga clic en un área en blanco de la pantalla para colocar el cursor con retícula roja y teclee dobles comillas ["]. Ahora usted puede teclear lo que quiera, como en un procesador Word. Al crecer la región, aparece una caja negra alrededor del texto. La caja tiene "manijas" de redimensionamiento en las orillas derecha e inferior del rectángulo. Una vez que haya terminado, haga clic fuera de la región de texto para volver a la introducción de operaciones matemáticas. La caja negra de selección desaparece cuando usted ya no está trabajando en la región de texto.

Operaciones matemáticas

Tipo	Vea en pantalla
1+	1 + ▪

Haga clic en algún punto de la esquina superior izquierda de la hoja de cálculo y la retícula roja se debe mover al lugar donde usted hizo clic. Después de teclear el número 1 y el signo + verá una cajita negra delimitada por líneas de edición azules. En Mathcad, esta caja negra se llama "marcador de posición" ("placeholder"). Si usted sigue tecleando, lo siguiente que escriba aparecerá en el marcador de posición. Por ejemplo, teclee 2 en el marcador de posición y luego teclee el signo igual (=) para ver el resultado.

$$1 + 2 = 3$$

En seguida se da una lista de las operaciones aritméticas básicas, junto con sus teclas y los equivalentes de botones en la **Pallette Calculator**.

Operación	Tecla	Pallette	Ejemplo
Suma	+	+	$2 + 2 = 4$
Resta	–	–	$2 - 2 = 0$
Multiplicación	*	×	$2 \cdot 2 = 4$
División	/	÷	$\frac{2}{2} = 1$
Exponenciación	^	x^Y	$2^2 = 4$

Observe que las operaciones aparecen en la hoja de cálculo Mathcad con la notación familiar: la multiplicación es un punto, la división una barra de fracción, los exponentes están en posición elevada, etc. Los cálculos se realizan internamente hasta 15 posiciones, pero usted puede mostrar menos posiciones en la respuesta. Para cambiar la forma predeterminada en la que se muestran los resultados numéricos y simbólicos en una hoja de cálculo, haga clic en un área en blanco de la hoja de cálculo. En seguida seleccione **Result** del menú **Format** para que aparezca la caja de diálogo **Result Format** y elija su configuración predeterminada. Asegúrese de marcar el botón rotulado "*Set as default*" y haga clic en OK. Si sólo quiere cambiar la forma en la que se muestra un resultado en particular, haga clic en la ecuación y siga los mismos pasos.

Aquí se dan algunos ejemplos adicionales que demuestran las características de Mathcad.

$$\sqrt{\frac{1.837 \cdot 10^3}{100 + 3^5}} = 2.3142353232$$

La mayoría de las funciones de ingeniería y de matemáticas están interconstruidas.

$$\log(1347.2) \cdot \text{sen}\left(\frac{3}{5} \cdot \pi\right) = 2.976$$

Las funciones y los operadores de Mathcad manejan fácilmente números complejos.

$$(2.3 + 4.7\text{i})^3 + \text{e}^{3-2\text{i}} = -148.613 - 47.498\text{i}$$

FUNCIONES Y VARIABLES MATEMÁTICAS

El símbolo de definición := se usa para definir una función o una variable en Mathcad. Por ejemplo, haga clic en una hoja de cálculo vacía para posicionar la retícula roja en un área en blanco y teclee:

Teclee	Verá en pantalla
f(x):x^2	$f(x) := x^2$

El símbolo de definición también se ubica en la selección **Evaluation** de la **Pallette Math**. Cuando usted cambia una función de definición o una variable, Mathcad recalcula inmediatamente los nuevos valores que dependan de ella. Una vez que usted ha definido una función como f(x), puede verla de varias maneras; por ejemplo,

$$f(x) := x^2$$

Ahora puede insertar un valor numérico como el argumento de f(x)

$$f(10) = 100$$

o definir una variable e insertarla como el argumento de f(x).

$$x := 3$$
$$f(x) = 9$$

Incluso puede definir otra función en términos de f(x):

$$g(y) := f(y) + 6$$
$$g(x) = 15$$

Observe que puede definir una función usando expresiones que construye a partir del teclado o de las paletas de operadores matemáticos. Usted también puede incluir cualquiera entre los cientos de funciones incorporadas de Mathcad. Para ver una lista de funciones incorporadas junto con breves descripciones, seleccione **Function** en el menú **Insert**, o haga clic en el botón ***f*(*x*)**. También

puede teclear el nombre de cualquier función interconstruida directamente en el teclado. Los siguientes son sólo unos pocos ejemplos que usan algunas de las funciones incorporadas de Mathcad:

Trig and Logs

$\ln(26) = 3.258 \qquad \csc(45 \cdot.\text{deg}) = 1.414$

Matrix Functions

$$\text{identity}(3) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Probability Distributions

$\text{pnorm}(2,0,1) = 0.977$

Variables de rango

En Mathcad, usted verá que desea trabajar con un rango de valores para muchas aplicaciones; por ejemplo, para definir una serie de valores para graficar. Por lo tanto, Mathcad proporciona el operador de rango (..), que se puede introducir tecleando un punto y coma (;). El primer número y el último establecen los puntos finales de la variable de rango, y el segundo define el incremento. Por ejemplo:

Teclear	Verá en pantalla
z:0,0.5;2	z:=0,0.5..2
z =	z = 0, 0.5, 1, 1.5, 2

Cálculos y operaciones matriciales

Para introducir una matriz, haga clic en el icono de matriz 3 × 3 en la **Pallette Matrix** (o elija **Matrix** en el menú **Insert**), elija el número de filas y columnas y luego llene los marcadores de lugar. Por ejemplo.

$$A := \begin{bmatrix} 4 & 5 & 1 \\ 5 & 0 & -12 \\ -7 & 2 & 8 \end{bmatrix}$$

Para calcular el inverso,

Teclear	Verá en pantalla
Aˆ–1=	$A^{-1} = \begin{bmatrix} 0.074 & -0.117 & -0.184 \\ 0.135 & 0.12 & 0.163 \\ 0.031 & -0.132 & -0.077 \end{bmatrix}$

Mathcad tiene un amplio conjunto de comandos para realizar diversas operaciones con matrices. Por ejemplo, para encontrar un determinante, teclee la barra vertical (|) o use el botón en la **Pallette Matrix**.

$|A| = 326$

Unidades

Mathcad puede también manejar unidades. Para ver las unidades interconstruidas, elija **Unit** en el menú **Insert**, o haga clic en el botón adecuado de la barra de herramientas. Para comenzar con un ejemplo sencillo, abra una nueva hoja de cálculo en Mathcad y teclee `Mass:75kg`[*Enter*]. Verá:

$\text{Mass} := 75 \text{ kg}$

También podría haber tecleado `Mass:75*kg`[*Enter*], multiplicando la cantidad por la unidad. Ahora, introduzca `g:9.8m/s^2` y deberá obtener

$$g := 9.8\frac{\text{m}}{\text{s}^2}$$

Para ver cómo Mathcad administra las unidades con los cálculos, introduzca `Mass*g=` y el resultado se debe mostrar con unidades combinadas como

$\text{Mass} \cdot g = 735 \text{ N}$

Mathcad usa el sistema de unidades SI como predeterminado, pero usted puede cambiar esto en **Tools**, **Worksheet Options**, **Unit System**. Los sistemas alternos incluyen CGS, MKS y US. En vez de teclear la unidad, puede también insertarla de una lista. Ensaye lo siguiente: Teclee `Temp:273.16*` y luego haga clic en **Insert, Unit**. Seleccione **Temperature** en la caja superior y **Kelvin(K)** en la caja inferior y haga clic en OK.

FUNCIÓN DE MÉTODOS NUMÉRICOS

Mathcad tiene varias funciones especiales interconstruidas que realizan diversas operaciones numéricas de interés especial para los lectores de este libro. En el texto se describen en detalle ejemplos de desarrollo y aplicación de estas funciones. Aquí se da una breve lista de algunas de las más importantes sólo para darle una visión general de las capacidades. Su uso se ilustra en las secciones correspondientes de este libro.

Nombre de la función	Uso
root	Resuelve f(x) = 0
polyroots	Encuentra todas las raíces de un polinomio
find	Resuelve un sistema de ecuaciones algebraicas no lineales
minerr	Devuelve una solución de error mínimo de un sistema de ecuaciones
lsolve	Resuelve un sistema de ecuaciones algebraicas lineales
linterp	Interpolación lineal
cspline	Interpolación cúbica de trazador (spline)
regress	Regresión polinómica
genfit	Regresión general no lineal
fft	Transformada de Fourier
ifft	Transformada de Fourier inversa
rkfixed	Resuelve un sistema de ecuaciones diferenciales usando un método de Runge-Kutta de cuarto orden de tamaño de paso fijo
rkadapt	Resuelve un sistema de ecuaciones diferenciales usando un método de Runge-Kutta de cuarto orden de tamaño de paso variable
sbval	Resuelve un problema de valor de frontera de dos puntos
eigenvals	Encuentra valores propios
eigenvecs	Encuentra vectores propios
relax	Resuelve ecuación de Poisson para un dominio cuadrado

PROCEDIMIENTOS Y SUBPROGRAMAS DE LÍNEAS MÚLTIPLES

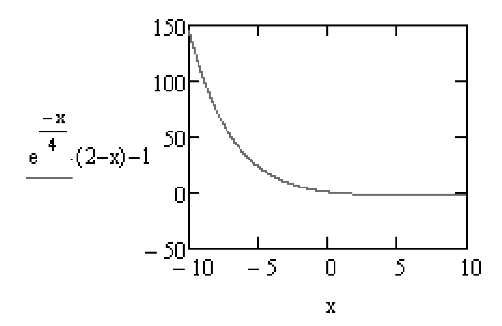
La **Pallette Programming** de Mathcad proporciona la capacidad para procedimientos o subprogramas de líneas múltiples con estructuras estándar de control como ciclos FOR y WHILE, ramificación, recursividad y más. Se pueden integrar subprogramas con las hojas de cálculo de Mathcad y pueden operar en escalares, vectores, matrices e incluso matrices de matrices.

CREACIÓN DE GRÁFICAS

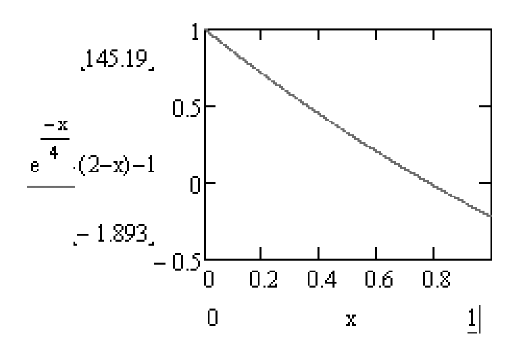
Las capacidades de Mathcad para gráficas son especialmente importantes para el trabajo de ingeniería. El primer tipo de gráfica de la que se debe saber es el *QuickPlot*. Comience con una nueva hoja de cálculo e introduzca la siguiente fórmula. Deje seleccionada la fórmula

$$e^{\frac{-x}{4}} \cdot (2 - x) - \underline{1}|$$

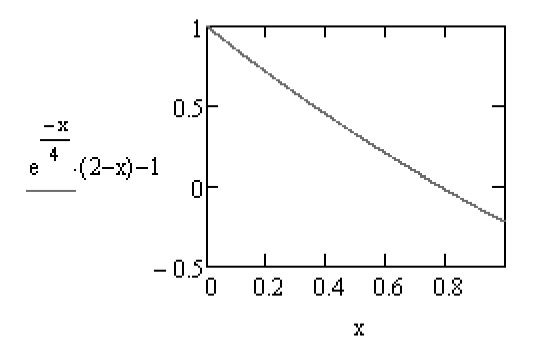
En el menú, seleccione **Insert**, **Graph**, **X-Y Plot**. Haga clic lejos de la gráfica y debe ver un *QuickPlot* automático.



Esta gráfica se puede ahora ajustar como se desee. Por ejemplo, haga clic en la gráfica para seleccionarla y cambie los límites de *x* a 0 a 1. Esto debe aparecer como



Luego haga clic lejos de la gráfica y el eje *y* se reescalará automáticamente. Haga doble clic en la gráfica y debe aparecer la caja de diálogo **Formatting Currently Selected X-Y Plot**. Revise las cajas para encontrar los ejes de coordenadas X y Y y haga clic en OK. Haga clic lejos de la gráfica y debe tener ahora



Mathcad grafica un punto para cada valor de la variable *x* del rango. Esta variable se creó automáticamente en este caso. Los puntos *x*-*y* se unen mediante segmentos cortos de línea recta. Usted puede crear su propia variable del rango para el eje *x*.

En vez de usar una fórmula para el eje *y*, usted puede usar una función. Para ilustrar esto, introduzca la siguiente definición de función *arriba* de su gráfica

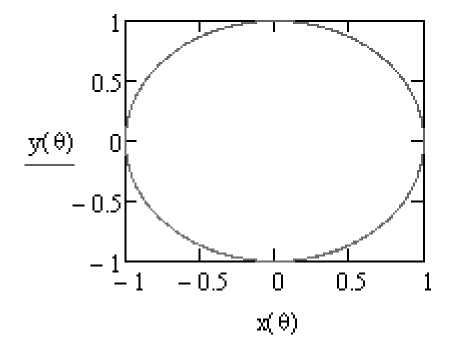
$$f(x) := e^{\frac{-x}{4}} \cdot (2 - x) - 1$$

y cambie el eje *y* de la fórmula a **f(x)**. Usted debe obtener la misma gráfica, pero ahora es en términos de la función **f(x)** que se usó para el eje *x*. Pruebe otro ejemplo debajo de su gráfica actual. Haga las siguientes definiciones:

$$N := 100 \qquad \theta := 0, \frac{2 \cdot x}{N} .. 2 \cdot x$$

$$x(\theta) := \cos(\theta) \qquad y(\theta) := \sin(\theta)$$

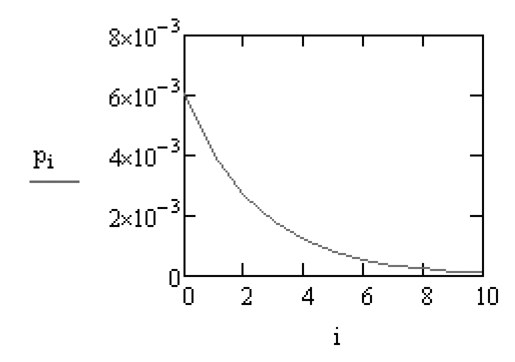
Inserte una gráfica en blanco oprimiendo la tecla @ (*shift-2*). Introduzca y(θ) en el marcador de lugar del eje *y*, y x(θ) en el marcador de lugar del eje *x*, y haga clic lejos de la gráfica. Esto debe dar la gráfica de una circunferencia.



También es posible graficar los elementos de un vector. Cree el siguiente vector de probabilidades binomiales:

$i := 0..\ 10$
$p_i := 0.4^i . 0.6^{(10\ -\ i)}$

e inserte un gráfico con **i** en el eje *x* y $\mathbf{p}_i$ en el eje *y* (puede usar la tecla [para el subíndice). Su gráfica debe verse como



Por supuesto, usted puede graficar un vector también contra otro, siempre y cuando ambos tengan el mismo número de elementos. Los vectores podrían contener datos en vez de una fórmula matemática.

Hay muchas configuraciones de ejes que se pueden ajustar haciendo doble clic en la gráfica. Éstas se explican por sí mismas en gran medida, y usted se puede familiarizar con ellas por la práctica.

Hay muchos otros estilos de gráficas que se pueden generar por Mathcad. Éstos incluyen gráficas polares, de superficie, de contorno, de barra 3D, dispersa, de campo vectorial y animaciones gráficas.

MATEMÁTICAS SIMBÓLICAS

Una característica intrigante y valiosa de Mathcad es su capacidad de llevar a cabo manipulaciones de matemática simbólica. Las capacidades simbólicas incluyen

- Manipulaciones algebraicas.
- Cálculo: diferenciación e integración.
- Resolución de ecuaciones algebraicas y sistemas de tales ecuaciones.

y más características avanzadas

- Transformadas simbólicas de Fourier, de Laplace y z.
- Optimización simbólica.

En seguida se reseña el primer grupo.

Se comenzará con un sencillo ejemplo de álgebra simbólica. Introduzca la expresión

$$(x + 2) \cdot (x - 1) \cdot (x + 4)$$

Expanda la línea horizontal de edición para que abarque toda la expresión y luego seleccione *Expand* del menú **Symbolics**. Usted debe ver

$$x^3 + 5 \cdot x^2 + 2 \cdot x - 8$$

Ahora, introduzca el polinomio $x^3 + 3x^2 + 3x + 1$ y siga el mismo procedimiento, pero elija **Factor** del menú **Symbolics**. Usted debe tener

$$x^3 + 3 \cdot x^2 + 3 \cdot x + 1$$
$$(x + 1)^3$$

Otra forma de llevar a cabo los comandos simbólicos es mediante teclas. Introduzca la siguiente expresión:

$$\boxed{\frac{x^2 - 3 \cdot x - 4}{x - 4} + 2 \cdot x - \underline{5}|}$$

Luego oprima la combinación de teclas ***Ctrl-Shift-***. y teclee `simplify` en el marcador de lugar que aparece. Usted debe obtener

$$\frac{x^2 - 3 \cdot x - 4}{x - 4} + 2 \cdot x - 5 \text{ simplify} \rightarrow 3 \cdot x - 4$$

Éste es un estilo diferente de evaluación simbólica que tiene la palabra clave retenida y muestra el resultado a la derecha. Usted ha visto el uso de tres importantes operadores simbólicos:

- expand Expandir todas las potencias y productos de las sumas.
- factor Factor dentro de un producto de funciones más simples.
- simplify Simplifica realizando aritmética, cancelando factores comunes, usando identidades y simplificando potencias.

Las características algebraicas adicionales incluyen expansión a una serie, expansión de fracciones parciales y extracción de coeficientes de un polinomio en un vector.

Ahora se va a experimentar con diferenciación simple. Introduzca la expresión

$$e^{\frac{-E}{R \cdot T}}$$

y deje seleccionada la T (o haga clic en ella para seleccionarla). Luego seleccione **Variable** y **Differentiate** del menú **Symbolics**. Usted debe tener el resultado

$$e^{\frac{-E}{R \cdot T}}$$

$$\frac{E \cdot e^{-\frac{E}{R \cdot T}}}{R \cdot T^2}$$

Otra forma de hacer esto es usar el operador de diferenciación de la barra de herramientas **Calculus**. Introduzca la siguiente expresión

$$\frac{d^2}{dx^2}\left(x^2 \cdot \sin(x)\right)$$

y luego haga clic en el botón → en la barra de herramientas **Evaluation**. (También puede oprimir **Ctrl-**.) El resultado debe ser

$$\frac{d^2}{dx^2}\left(x^2 \cdot \sin(x)\right) \rightarrow 2 \cdot \sin(x) - x^2 \cdot \sin(x) + 4 \cdot x \cdot \cos(x)$$

Las integrales simbólicas se pueden determinar en forma ya sea definida o indefinida. Para una integral indefinida, comience tecleando ***Ctrl-i*** o haga clic en el botón ∫ en la barra de herramientas **Calculus**. Luego introduzca la función deseada y la diferencial, seguidas del botón → para producir el siguiente resultado:

$$\int e^{-x}\,dx \rightarrow -e^{-x}$$

Mathcad puede también calcular límites simbólicamente. Los botones adecuados están en la barra de herramientas **Calculus**.

PARA APRENDER MÁS ACERCA DE MATHCAD

En esta breve introducción, se ha cubierto sólo lo básico de Mathcad. Hay disponible asistencia adicional en el paquete de software Mathcad en diversas formas.

ToolTips

Deje flotar el cursor del ratón sobre un botón de paleta o barra de herramientas durante unos pocos segundos. Verá una nota explicativa cerca del botón. Vea también en la línea de mensaje en la parte inferior de la ventana de aplicación Mathcad para encontrar pistas y atajos útiles.

Centro de recursos y QuickSheets

Para ayudarle a comenzar rápido y mantenerle aprendiendo, Mathcad viene completo con *QuickSheets*. Éstas proporcionan atajos matemáticos para análisis de uso frecuente —desde graficar una función hasta resolver ecuaciones simultáneas y realizar análisis de varianza—. Hay numerosas *QuickSheets*. Para abrir la sección de *QuickSheets*, elija **QuickSheets** de la opción **Help** en el menú **Main**.

Ayuda en línea

Online Help proporciona instrucciones detalladas, paso a paso, para usar todas las características de Mathcad. Siempre hay ayuda disponible en cualquier momento simplemente acudiendo al botón **Help** en el menú **Main**. Allí encontrará varios vínculos, incluido el sitio de red Mathcad y Mathcad Training.

BIBLIOGRAFÍA

Al-Khafaji, A. W. y J. R. Tooley, *Numerical Methods in Engineering Practice,* Holt, Rinehart y Winston, Nueva York, 1986.

Allaire, P. E., *Basics of the Finite Element Method,* William C. Brown, Dubuque, 1985.

Ames, W. F., *Numerical Methods for Partial Differential Equations,* Academic Press, Nueva York, 1977.

Ang, A. H-S. y W. H. Tang, *Probability Concepts in Engineering Planning and Design, vol. 1: Basic Principles,* Wiley, Nueva York, 1975.

APHA (American Public Health Association). *Standard Methods for the Examination of Water and Wastewater*, 18a. ed., Washington, DC, 1992.

Atkinson, K. E., *An Introduction to Numerical Analysis,* Wiley, Nueva York, 1978.

Atkinson, L. V. y P. J. Harley, *An Introduction to Numerical Methods with Pascal,* Addison-Wesley, Reading, MA, 1983.

Baker, A. J., *Finite Element Computational Fluid Mechanics,* McGraw-Hill, Nueva York, 1983.

Bathe, K.-J. y E. L. Wilson, *Numerical Methods in Finite Element Analysis,* Prentice-Hall, Englewood Cliffs, 1976.

Booth, G. W. y T. L. Peterson, "Nonlinear Estimation", I.B.M. Share Program Pa. núm. 687 WLNL1, 1958.

Boyce, W. E. y R. C. DiPrima, *Elementary Differential Equations and Boundary Value Problems,* 5a. ed. Wiley, Nueva York, 1992.

Branscomb, L. M., "Electronics and Computers: An Overview", *Science,* 215: 755, 1982.

Brent, R. P., *Algorithms for Minimization Without Derivatives*, Prentice-Hall, Englewood Cliffs, NJ, 1973.

Brigham, E. O., *The Fast Fourier Transform,* Prentice-Hall, Englewood Cliffs, NJ, 1974.

Burden, R. L. y J. D. Faires, *Numerical Analysis,* 8a. ed., PWS Publishing, Boston, 2005.

Butcher, J. C., "On Runge-Kutta Processes of Higher Order", *J. Austral. Math. Soc.,* 4: 179, 1964.

Carnahan, B., H. A. Luther y J. O. Wilkes, *Applied Numerical Methods,* Wiley, Nueva York, 1969.

Cash, J. R. y A. H. Karp, *ACM Transactions on Mathematical Software,* 16: 201-222, 1990.

Chapra, S. C., *Surface Water-Quality Modeling,* McGraw-Hill, Nueva York, 1997.

Chapra, S. C., *Applied Numerical Methods with MATLAB,* 2a. ed., McGraw-Hill, Nueva York, N.Y., 2007.

Cheney, W. y D. Kincaid, *Numerical Mathematics and Computing,* 6a. ed., Brooks/Cole, Monterey, CA, 2008.

Chirlian, P. M., *Basic Network Theory,* McGraw-Hill, Nueva York, 1969.

Cooley, J. W., P. A. W. Lewis y P. D. Welch, "Historical Notes on the Fast Fourier Transform", *IEEE Trans. Audio Electroacoust.,* AU-15(2): 76-79, 1977.

Dantzig, G. B., *Linear Programming and Extensions,* Princeton University Press, Princeton, NJ, 1963.

Davis, H. T., *Introduction to Nonlinear Differential and Integral Equations,* Dover, Nueva York, 1962.

Davis, L., *Handbook of Genetic Algorithms,* Van Nostrand Reinhold, Nueva York, 1991.

Davis, P. J. y P. Rabinowitz, *Methods of Numerical Integration,* Academic Press, Nueva York, 1975.

Dennis, J. E. y R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations,* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996.

Dekker, T. J., *Finding a zero by means of successive linear interpolation,* in Constructive Aspects of the Fundamental Theorem of Algebra, B. Dejon y P. Henrici (eds.), Wiley-Interscience, Nueva York, 1969, pp. 37-48.

Dijkstra, E. W., "Go To Statement Considered Harmful", *Commun. ACM,* 11(3): 147-148, 1968.

Draper, N. R. y H. Smith, *Applied Regression Analysis,* 2a. ed., Wiley, Nueva York, 1981.

Enright, W. H., T. E. Hull y B. Lindberg, "Comparing Numerical Methods for Stiff Systems of ODE's", *BIT,* 15: 10, 1975.

Fadeev, D. K. y V. N. Fadeeva, *Computational Methods of Linear Algebra,* Freeman, San Francisco, 1963.

Ferziger, J. H., *Numerical Methods for Engineering Application,* Wiley, Nueva York, 1981.

Fletcher, R., *Practical Methods of Optimization: 1: Unconstrained Optimization,* Wiley, Chichester, UK, 1980.

Fletcher, R., *Practical Methods of Optimization: 2: Constrained Optimization,* Wiley, Chichester, 1981.

Forsythe, G. E. y W. R. Wasow, *Finite-Difference Methods for Partial Differential Equations,* Wiley, Nueva York, 1960.

Forsythe, G. E., M. A. Malcolm y C. B. Moler, *Computer Methods for Mathematical Computation,* Prentice-Hall, Englewood Cliffs, NJ, 1977.

Fylstra, D., L. S. Lasdon, J. Watson y A. Waren, "Design and Use of the Microsoft Excel Solver", *Interfaces,* 28(5): 29-55, 1998.

Gabel, R. A. y R. A. Roberts, *Signals and Linear Systems,* Wiley, Nueva York, 1987.

Gear, C. W., *Numerical Initial-Value Problems in Ordinary Differential Equations,* Prentice-Hall, Englewood Cliffs, NJ, 1971.

Gerald, C. F. y P. O. Wheatley, *Applied Numerical Analysis,* 7a. ed., Addison-Wesley, Reading, MA, 2004.

Gill, P. E., W. Murray y M. H. Wright, *Practical Optimization,* Academic Press, Londres, 1981.

Gladwell, J. y R. Wait, *A Survey of Numerical Methods of Partial Differential Equations,* Oxford University Press, Nueva York, 1979.

Goldberg, D. E., *Genetic Algorithms in Search, Optimization and Machine Learning,* Addison-Wesley, Reading, MA, 1989.

Guest, P. G., *Numerical Methods of Curve Fitting,* Cambridge University Press, Nueva York, 1961.

Hamming, R. W., *Numerical Methods for Scientists and Engineers,* 2a. ed., McGraw-Hill, Nueva York, 1973.

Hartley, H. O., "The Modified Gauss-Newton Method for Fitting Nonlinear Regression Functions by Least Squares", *Technometrics* 3: 269-280, 1961.

Hayt, W. H. y J. E. Kemmerly, *Engineering Circuit Analysis,* McGraw-Hill, Nueva York, 1986.

Heideman, M. T., D. H. Johnson y C. S. Burrus, "Gauss and the History of the Fast Fourier Transform", *IEEE ASSP Mag.,*1(4): 14-21, 1984.

Henrici, P. H., *Elements of Numerical Analysis,* Wiley, Nueva York, 1964.

Hildebrand, F. B., *Introduction to Numerical Analysis,* 2a. ed., McGraw-Hill, Nueva York, 1974.

Hillier, F. S. y G. J. Lieberman, *Introduction to Operations Research*, 8a. ed., McGraw-Hill, Nueva York, 2005.

Hoffman, J., *Numerical Methods for Engineers and Scientists,* McGraw-Hill, Nueva York, 1992.

Holland, J. H., *Adaptation in Natural and Artificial Systems,* University of Michigan Press, Ann Arbor, MI, 1975.

Hornbeck, R. W., *Numerical Methods,* Quantum, Nueva York, 1975.

Householder, A. S., *Principles of Numerical Analysis,* McGraw-Hill, Nueva York, 1953.

Householder, A. S., *The Theory of Matrices in Numerical Analysis,* Blaisdell, Nueva York, 1964.

Householder, A. S., *The Numerical Treatment of a Single Nonlinear Equation,* McGraw-Hill, Nueva York, 1970.

Huebner, K. H. y E. A. Thornton, *The Finite Element Method for Engineers,* Wiley, Nueva York, 1982.

Hull, T. E. y A. L. Creemer, "The Efficiency of Predictor-Corrector Procedures", *J. Assoc. Comput. Mach.,* 10: 291, 1963.

Isaacson, E. y H. B. Keller, *Analysis of Numerical Methods,* Wiley, Nueva York, 1966.

Jacobs, D. (ed.), *The State of the Art in Numerical Analysis,* Academic Press, Londres, 1977.

James, M. L., G. M. Smith y J. C. Wolford, *Applied Numerical Methods for Digital Computations with FORTRAN and CSMP,* 3a. ed., Harper & Row, Nueva York, 1985.

Keller, H. B., *Numerical Methods for Two-Point Boundary-Value Problems,* Wiley, Nueva York, 1968.

Lapidus, L. y G. F. Pinder, *Numerical Solution of Partial Differential Equations in Science and Engineering,* Wiley, Nueva York, 1981.

Lapidus, L. y J. H. Seinfield, *Numerical Solution of Ordinary Differential Equations,* Academic Press, Nueva York, 1971.

Lapin, L. L., *Probability and Statistics for Modern Engineering,* Brooks/Cole, Monterey, CA, 1983.

Lasdon, L. S. y S. Smith, "Solving Large Nonlinear Programs Using GRG", *ORSA Journal on Computing,* 4(1): 2-15, 1992.

Lasdon, L. S., A. Waren, A. Jain y M. Ratner, "Design and Testing of a Generalized Reduced Gradient Code for Nonlinear Programming", *ACM Transactions on Mathematical Software,* 4(1): 34-50, 1978.

Lawson, C. L. y R. J. Hanson, *Solving Least Squares Problems,* Prentice-Hall, Englewood Cliffs, NJ, 1974.

Luenberger, D. G., *Introduction to Linear and Nonlinear Programming,* Addison-Wesley, Reading, MA, 1984.

Lyness, J. M., "Notes on the Adaptive Simpson Quadrature Routine", *J. Assoc. Comput. Mach.,* 16: 483, 1969.

Malcolm, M. A. y R. B. Simpson, "Local Versus Global Strategies for Adaptive Quadrature", *ACM Trans. Math. Software,* 1: 129, 1975.

Maron, M. J., *Numerical Analysis, A Practical Approach,* Macmillan, Nueva York, 1982.

Milton, J. S. y J. C. Arnold, *Introduction to Probability y Statistics: Principles and Applications for Engineering and the Computing Sciences,* 4a. ed., McGraw-Hill, Nueva York, 2002.

Moler, C. B., *Numerical Computing with MATLAB,* SIAM, Philadelphia, 2004.

Muller, D. E., "A Method for Solving Algebraic Equations Using a Digital Computer", *Math. Tables Aids Comput.,* 10: 205, 1956.

Na, T. Y., *Computational Methods in Engineering Boundary Value Problems,* Academic Press, Nueva York, 1979.

Noyce, R. N., "Microelectronics", *Sci. Am.* 237: 62, 1977.

Oppenheim, A. V. y R. Schafer, *Digital Signal Processing,* Prentice-Hall, Englewood Cliffs, NJ, 1975.

Ortega, J. y W. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables,* Academic Press, Nueva York, 1970.

Ortega, J. M., *Numerical Analysis—A Second Course,* Academic Press, Nueva York, 1972.

Prenter, P. M., *Splices y Variational Methods,* Wiley, Nueva York, 1975.

Press, W. H., B. P. Flanner, S. A. Teukolsky y W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing,* 3a. ed. Cambridge University Press, Cambridge, 2007.

Rabinowitz, P., "Applications of Linear Programming to Numerical Analysis", *SIAM Rev.,* 10: 121-159, 1968.

Ralston, A., "Runge-Kutta Methods with Minimum Error Bounds", *Match. Comp.,* 16: 431, 1962.

Ralston, A. y P. Rabinowitz, *A First Course in Numerical Analysis,* 2a. ed., McGraw-Hill, Nueva York, 1978.

Ramirez, R. W., *The FFT, Fundamentals and Concepts,* Prentice-Hall, Englewood Cliffs, NJ, 1985.

Rao, S. S., *Engineering Optimization: Theory and Practice,* 3a. ed., Wiley-Interscience, Nueva York, 1996.

Revelle, C. S., E. E. Whitlach y J. R. Wright, *Civil and Environmental Systems Engineering,* Prentice-Hall, Englewood Cliffs, NJ, 1997.

Rice, J. R., *Numerical Methods, Software and Analysis,* McGraw-Hill, Nueva York, 1983.

Ruckdeschel, F. R., *BASIC Scientific Subroutine, vol.* 2, Byte/McGraw-Hill, Peterborough, NH, 1981.

Scarborough, J. B., *Numerical Mathematical Analysis,* 6a. ed., Johns Hopkins Press, Baltimore, MD, 1966.

Scott, M. R. y H. A. Watts, "A Systematized Collection of Codes for Solving Two-Point Boundary-Value Problems", en *Numerical Methods for Differential Equations,* L. Lapidus y W. E. Schiesser (eds.), Academic Press, Nueva York, 1976.

Shampine, L. F. y R. C. Allen, Jr., *Numerical Computing: An Introduction,* Saunders, Philadelphia, 1973.

Shampine, L. F. y C. W. Gear, "A User's View of Solving Stiff Ordinary Differential Equations", *SIAM Review,* 21: 1, 1979.

Simmons, E. F., *Calculus with Analytical Geometry,* McGraw-Hill, Nueva York, 1985.

Stark, P. A., *Introduction to Numerical Methods,* Macmillan, Nueva York, 1970.

Stasa, F. L., *Applied Finite Element Analysis for Engineers,* Holt, Rinehart and Winston, Nueva York, 1985.

Stewart, G. W., *Introduction to Matrix Computations,* Academic Press, Nueva York, 1973.

Swokowski, E. W., *Calculus with Analytical Geometry,* 2a. ed., Prindle, Weber and Schmidt, Boston, 1979.

Taylor, J. R., *An Introduction to Error Analysis,* University Science Books, Mill Valley, CA, 1982.

Tewarson, R. P., *Sparse Matrices,* Academic Press, Nueva York, 1973.

Thomas, G. B., Jr. y R. L. Finney, *Calculus and Analytical Geometry,* 5a. ed., Addison-Wesley, Reading, MA, 1979.

Van Valkenburg, M. E., *Network Analysis,* Prentice-Hall, Englewood Cliffs, NJ, 1974.

Varga, R., *Matrix Iterative Analysis,* Prentice-Hall, Englewood Cliffs, NJ, 1962.

Vichnevetsky, R., *Computer Methods for Partial Differential Equations, vol. 1: Elliptical Equations and the Finite Element Method,* Prentice-Hall, Englewood Cliffs, NJ, 1981.

Vichnevetsky, R., *Computer Methods for Partial Differential Equations, vol. 2: Initial Value Problems,* Prentice-Hall, Englewood Cliffs, NJ, 1982.

Wilkinson, J. H., *The Algebraic Eigenvalue Problem,* Oxford University Press, Fair Lawn, NJ, 1965.

Wilkinson, J. H. y C. Reinsch, *Linear Algebra: Handbook for Automatic Computation,* vol. 11, Springer-Verlag, Berlin, 1971.

Wold, S., "Spline Functions in Data Analysis", *Technometrics,* 16(1): 1-11, 1974.

Yakowitz, S. y F. Szidarovsky, *An Introduction to Numerical Computation,* Macmillan, Nueva York, 1986.

Young, D. M., *Iterative Solution of Large Linear Systems,* Academic Press, Nueva York, 1971.

Zienkiewicz, O. C., *The Finite Element Method in Engineering Science,* McGraw-Hill, Londres, 1971.

ÍNDICE ANALÍTICO

La n después de un número de página indica que la referencia se encuentra en las notas.

D

E

F

G

H

I

L

M

N

O

T

La séptima edición de *Métodos numéricos para ingenieros* continúa ofreciendo una presentación innovadora y accesible sobre una amplia gama de métodos numéricos. Dado que regularmente se emplea software para el análisis numérico, esta revisión mantiene un fuerte enfoque en el uso apropiado de las herramientas de cómputo, así como de las discusiones de los fundamentos matemáticos subyacentes.

Las características principales de esta edición son:

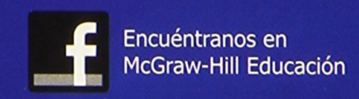
- Numerosos problemas que han sido tomados de situaciones reales de la práctica de la ingeniería; muchos, incluso, se retoman de áreas nuevas como la bioingeniería.
- Cobertura amplia de los métodos numéricos, que incluye la revisión de temas como optimización y ecuaciones diferenciales.
- Enfoque en las herramientas de cómputo, que considera algoritmos de pseudocódigo y softwares populares como MATLAB, Excel y MathCAD.
- Ejemplos de excelencia y estudios de caso que cubren todas las disciplinas de la ingeniería; los estudiantes podrán emplear estas habilidades en el campo profesional que elijan.











Visite nuestra página web
www.mcgraw-hill-educacion.com